# LINGUAGENS ESPECIALIZADAS

# EM *CORPORA*

## MODOS DE DIZER E INTERFACES DE PESQUISA

Pontifícia Universidade Católica
do Rio Grande do Sul

*Cristina Lopes Perna*

*Heloísa Koch Delgado*

*Maria José Finatto*

*(Organizadoras)*

# LINGUAGENS ESPECIALIZADAS EM *CORPORA*

## MODOS DE DIZER E INTERFACES DE PESQUISA

Porto Alegre
2010

**CAPA** Luciana Braun
**COLABORAÇÃO** Luiza Rico Bini
**REVISÃO DE TEXTO** Cristina Perna e Heloísa Delgado
**EDITORAÇÃO ELETRÔNICA** Gabriela Viale Pereira

**EDIPUCRS – Editora Universitária da PUCRS**
Av. Ipiranga, 6681 – Prédio 33
Caixa Postal 1429 – CEP 90619-900
Porto Alegre – RS – Brasil
Fone/fax: (51) 3320 3711
e-mail: edipucrs@pucrs.br - www.pucrs.br/edipucrs.

Dados Internacionais de Catalogação na Publicação (CIP)

L755 Linguagens especializadas em corpora : modos de dizer e interfaces de pesquisa [recurso eletrônico] / organizadoras, Cristina Lopes Perna, Heloísa Koch Delgado, Maria José Finatto. – Dados eletrônicos. – Porto Alegre : EDIPUCRS, 2010.
255 p.

Publicação Eletrônica
Sistema requerido: Adobe Acrobat Reader
Modo de Acesso: < http://www.pucrs.br/edipucrs/>
ISBN 978-85-397-0024-0 (on-line)

1. Linguística Computacional. 2. Processamento da Linguagem Natural. 3. Terminologia. 4. Linguística – Pesquisas. I. Perna, Cristina Lopes. II. Delgado, Heloísa Koch. III. Finatto, Maria José.

CDD 410

**Ficha Catalográfica elaborada pelo Setor de Tratamento da Informação da BC-PUCRS.**

# SUMÁRIO

# LINGUAGENS ESPECIALIZADAS E TERMINOLOGIA: O PASSADO PROJETANDO O FUTURO

*Anna Maria Becker Maciel*[1]

**Resumo:** *O objetivo deste texto é chamar a atenção para o estreito relacionamento da pesquisa das linguagens especializadas e a pesquisa terminológica. Para tanto, recorre-se àqueles que, no período áureo da tradição wüsteriana, realizaram pesquisa fundamental sobre a então chamada língua da ciência e da tecnologia à luz de pressupostos da teoria linguística. De acordo com seus ensinamentos, demonstra-se que a expressão linguística dos especialistas não se limita ao uso de uma dada terminologia; mas que se conforma ao padrão peculiar de comunicação característico de diferentes áreas do conhecimento. Tal padrão resulta da seleção de itens lexicais e da preferência de construções sintáticas, processos semânticos e propósitos pragmáticos, condicionados ao contexto de comunicação próprio da área especializada. Nessa perspectiva, a integração que existe entre a linguagem especializada e sua terminologia não permite que pesquisas sejam conduzidas separadamente.*

## 1. Introdução

O que é uma linguagem especializada? Existem linguagens especializadas sem terminologia? Existem terminologias fora das linguagens especializadas? As linguagens especializadas se compõem apenas de termos? É possível integrar o estudo científico das linguagens especializadas, em outras palavras, a linguística das linguagens especializadas e o estudo científico dos termos, isto é, a terminologia?

Essas são as questões que levanto e busco resposta através da releitura do legado dos linguistas que nos precederam na pesquisa da linguagem utilizada na ciência e na técnica e no estudo da terminologia

---

[1] Grupo de Pesquisa TERMISUL Projeto Terminológico Cone Sul – Programa de Pós-graduacao Letras - Universidade Federal do Rio Grande do Sul (UFRGS).

nos últimos cinquenta anos. Dentre esses, sem menosprezar outros investigadores de igual autoridade, nos limites deste trabalho, recorro a Rostilav Kocourek, Juan Carlos Sager, Pierre Lérat e Lothar Hoffmann. Esses mestres, contemporâneos nossos, trilharam longos caminhos desde o surgimento da Teoria Geral da Terminologia (TGT), atribuída a Eugen Wüster (1974). Cada um ao seu modo, no contexto acadêmico em que desenvolveu sua atividade, vivenciou a evolução da fase pioneira dos estudos terminológicos normativos exclusivamente centrados na denominação de conceitos para a fase contemporânea dos estudos descritivos que contemplam os termos em sua realização nos textos. Essa última fase se estende até nossos dias e se caracteriza pela efervescência de propostas teórico-metodológicas em que se observa a preocupação com o estudo do texto.

Meu objetivo é contribuir para dirimir dúvidas e mal-entendidos, remover obstáculos, vencer preconceitos e estimular o avanço da pesquisa das linguagens especializadas e da terminologia antes que acirrar discussões. Proponho uma reflexão desenvolvida em três momentos: passado, presente e futuro. Para tanto, faço uma breve retomada das obras em que os quatro mestres expõem seu pensamento sobre questões que ainda hoje são debatidas. Em seguida, procuro apresentar o que, no meu entendimento, parece ser hoje aceito, se não por todos, pelo menos por muitos dos que procuram resposta para a mesma problemática. Finalmente, sem ter a pretensão de trazer soluções, não sugiro uma volta ao passado, mas um esforço maior de aprofundamento de pressupostos fundamentais que levem à revisão de procedimentos metodológicos do presente com os olhos postos no futuro.

## 2. O passado

Na perspectiva da teoria da comunicação e sob a luz da teoria dos atos de fala, J. C. Sager, professor do MIST (*Manchester Institute of Science and Technology)* da Universidade de Manchester, com a colaboração de D.Dungworth e P.F. McDonald, publicou *English Special Languages: principles and practice in science and technology,* em 1980. Nessa obra, em uma abordagem essencialmente linguística,

os autores procuraram descrever a linguagem especializada em todos os seus aspectos: cognitivo, semântico, pragmático, sintático e morfológico. Não se afastaram da tendência então vigente de considerar as linguagens especializadas como sistemas de comunicação restritos aos especialistas, mas perceberam a realidade do uso da linguagem especializada em um universo multidisciplinar sem fronteiras bem delimitadas entre os vários campos de saber. Sem conceder prioridade ao enfoque cognitivo, não o ignoraram, enfatizaram os propósitos visados pelos interlocutores e focalizaram a expressão linguística da informação transmitida. Fruto da reflexão teórica e da prática de professores de língua inglesa que fundamentaram, na linguística, a experiência docente adquirida no manuseio constante de textos especializados, a obra se constitui em um verdadeiro tratado de linguística da língua inglesa, usada na ciência e na tecnologia.

Sua definição de línguas especiais, *special languages*, se tornou clássica: "sistemas semióticos complexos semi-autônomos, baseados e derivados da língua comum, seu uso pressupõe educação especial e é restrito à comunicação entre especialistas da mesma área e de áreas estreitamente relacionadas" (SAGER, DUNGWORTH E MACDONALD, 1980, p.69[2]; Minha tradução).

Hoje, Sager (2007, p. 116), diante das exigências de um mundo globalizado e da multiplicação das necessidades de informação do grande público, admite que essa definição merece ser, em parte, reformulada e que a análise e a descrição das linguagens especializadas devem ser revisadas, exceto no que diz respeito às disciplinas puramente científicas.

Pierre Lerat, professor de linguística da *Université Paris-Nord*, também enfatiza as funções da linguagem especializada e sua pertinência ao sistema da língua, mas prefere dizer *langues spécialisées,* línguas especializadas, título que escolheu para sua obra publicada em 1995. Nela, ele desenvolve um estudo pluridimensional

---

[2] *Special languages are semi-autonomous, complex, semiotic systems based and derived from general language; their use presupposes special education and is restricted to communication among specialists in the same and closely related fields. (*SAGER, DUNGWORTH, MACDONALD, 1980, p. 69)

centrado nos pressupostos da linguística geral, examinando a linguagem da ciência e da técnica sob o ponto de vista gramatical e semântico. Ele define uma língua especializada[3] como:

> [...] o uso de uma língua natural para dar conta tecnicamente de conhecimentos especializados [...]. A língua especializada é, em primeiro lugar, uma língua em situação de uso profissional (uma 'língua na especialidade' como diz a escola de Praga). É a própria língua como sistema autônomo, mas ao serviço de uma função maior: a transmissão de conhecimentos. (LERAT, 1995, p.221. Minha tradução).

Como Sager *et alii*, Lerat também enfatiza o aspecto pragmático da comunicação do conhecimento manifesto na linguagem especializada, mas enquanto aqueles procuraram fundamentação na teoria dos atos de fala de Austin (1978), ele busca apoio na teoria da enunciação de Benveniste (1989). Lerat afirma que seu livro é uma reflexão global sobre as "línguas de especialidade", raramente abordadas, com preocupações de teoria linguística, e defende que uma teoria das línguas especializadas só pode ser baseada em uma teoria geral das línguas que considere os termos como denominações de conhecimentos (LERAT, 1995, p.24).

Em 1982, Kocourek, linguista de origem eslava, professor de linguística do Departamento de Estudos Franceses da Universidade Dalhousie, em Halifax no Quebec, publicou *La Langue Française de la Technique et de la Science: vers une linguistique de la langue savante.* O subtítulo, "para uma linguística da língua erudita",[4] anuncia uma corajosa empreitada que ainda hoje está em pauta e é objeto de

---

[3] [...] *l'usage d'une langue naturelle pour rendre compte techniquement de connaissances spécialisées* [...] *La langue spécialisée est d'abord une langue en situation d'emploi professionnel (une 'langue en spécialité', comme dit l'école de Prague). C'est la langue ellle-même (comme système autonome) mais au service d'une fonction majeure: la transmission de connaissances.* (LÉRAT, 1995, p.21).

[4] *C'est la langue de la recherche philosophique, sociologique, pédagogique, linguistique, technoscientifique (au sens ci-dessus) géographique et historique (classes CDU)* (KOCOUREK ,1991, p.35).

controvertidos debates de terminólogos que somente consideram os termos e de linguistas que só acreditam nos textos.

Para Kocourek (1991, 36)[5] a língua de especialidade (*langue de spécialité*), que ele coloca como hiperônimo de língua tecnocientífica *(langue technoscientifique)* e língua erudita (*langue savante*),

> [...] é uma *variedade* de língua com predominância *cognitiva,*
> - cujos *textos*, cumulativos, de emotividade, subjetividade, metaforicidade controlados, e delimitados de maneira externa, têm por objetivo significar e comunicar, no seio de uma coletividade restrita, o conteúdo temático ponderado e circunstanciado, e
> - cujos recursos*,* que subjazem a esses textos sobre todos os planos linguísticos, são marcados por caracteres gráficos, por tendências sintáticas e sobretudo por um conjunto rapidamente renovável de unidades lexicais que exigem, e recebem nos textos uma precisão semântica metalinguística (p.42) (Minha tradução).

Kocourek enfatiza em sua obra que a língua especializada é um subconjunto da língua comum com recursos próprios e de uso restrito a uma coletividade que partilha os mesmos interesses. Tais recursos são os mesmos da língua comum, mas são selecionados, adaptados e enriquecidos de acordo com as necessidades e os propósitos da área, tendo em vista a preservação do conteúdo a ser transmitido. Nessa perspectiva, a linguagem de especialidade é antes de tudo um meio de expressão de um conteúdo temático.

---

[5] *La langue de spécialité est une variété de langue à dominante 'cognitive',*
- *dont les 'textes', cumulatifs, d'émotivité, de subjectivité, et de métaphoricité contrôlées, et délimités de manière externe, ont pour but de signifier et de communiquer, ao sein d'une collectivité restreinte, le contenu thématique, raisonné et circonstancié,*
- *et dont les resources, qui sous-tendent ces textes sur tous les plans linguistiques, par des tendances syntaxiques, et surtout, par um ensemble rapidement renouvelable des unités lexicales qui requièrent, et reçoivent dans les textes, une précision sémantique métalinguistique (ibidem, p.36)*

Esse conteúdo é transmitido por palavras que referem conceitos específicos e que constituem uma terminologia própria, que separada do seu meio de expressão natural, o texto especializado, não subsiste. Desse modo, o estudo da terminologia complementa o estudo dos textos especializados, diz Kocourek, (1991b, p.75)[6] e ainda acrescenta que dizer que a abordagem terminológica foi substituída pela linguística dos textos especializados não é exato. O estudo da linguagem especializada foi ampliado a fim de abranger o plano textual e que esse estudo hoje é mais complexo e satisfatório (minha tradução).

Alguns anos depois da primeira edição publicada em 1981, e em 1991, reeditada a obra, o autor ampliou a discussão sobre a necessidade de tal linguística. Nesse sentido, reforçou a proposta calcada sobre as características da língua de especialidade e o relacionamento fundamental entre texto e termo, que ele caracteriza como "dois fatores irredutíveis da linguagem especializada". Justamente esse tópico interessa sobremaneira à reflexão aqui empreendida.

Deve-se a Lothar Hoffmann e ao grupo de linguistas germânicos por ele encabeçados, a publicação de uma obra gigantesca de 1369 páginas, contendo uma coleção de artigos sobre a pesquisa das linguagens especializadas e da terminologia *Fachsprachen. Languages for Special Purposes, Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft An International Handbook of Special Language and Terminology Research*. ".[7] O que chama atenção nessa obra é que seu conteúdo reúne a pesquisa das linguagens especializadas e a terminologia. Essa combinação em um empreendimento de tanto fôlego me leva a acreditar que também o grupo germânico não quer afastar o estudo texto especializado da pesquisa de seus termos, corroborando aquilo que tanto Sager, Lerat e Kocourek afirmaram já nos anos 80 como consequência de seus

---

[6] *Dire que l'approche terminologique a eté remplacée par la linguistique des textes spécialisés n'est pas exact. C'est plutôt que l'étude de la langue spécialisée a été élargie afin de comprendre le plan textuel, et qu'elle est aujourd'hui plus complexe et satisfaisante* (Kocourek, (1991b, p.75)

[7] Minha tradução: Línguas para Finalidades Específicas. Um Manual Internacional para a Pesquisa da Linguagem Especializada e da Terminologia".

estudos nas décadas precedentes 80 e que o próprio Hoffmann também sempre defendeu.

Lothar Hoffmann, que, desde os anos sessenta até 1993, exerceu a docência de Linguística Aplicada na Universidade de Leipzig e ainda hoje mantém contacto com alunos e pesquisadores, também acredita na necessidade de um estudo linguístico sistemático das linguagens especializadas, fundamentado em sólidos princípios teórico-metodológicos. Suas numerosas obras, entre as quais destaco, *Kommunikationsmittel Fachsprache* (1967) e *Zum Fachwort zum Fachtext* (1988)[8] apresentam as linhas mestras da área de estudo que ele denomina *Fachsprachenlinguistik* (Linguística das Línguas Especializadas). Segundo ele, essa linguística dá especial atenção ao texto como um todo funcional e estrutural da língua comum que usa determinadas formas lexicais, construções sintáticas e textuais que caracterizam a comunicação da área temática. Ao mesmo tempo, valoriza a terminologia como o núcleo linguístico da linguagem especializada e salienta que a pesquisa das terminologias e das construções sintáticas do texto especializado continua necessária e produtiva (HOFFMANN 1988, *apud* BRUMME, 1998, p,71-78).

Nesse contexto, coloca em igual patamar a pesquisa do texto e a pesquisa das palavras nele usadas, afirmando que as unidades léxicas do texto especializado pertencem em sentido mais amplo ao léxico especializado, já que contribuem para o processo de comunicação de uma maneira direta ou indireta. Nesse entendimento, juntamente com as palavras tradicionalmente consideradas termos, as palavras funcionais, como conjunções, preposições e as palavras não tematicamente marcadas merecem ser descritas e analisadas.

Em consonância com Finatto, (2001, p.61), quero destacar que “quando se fala em texto especializado, não se pode deixar de mencionar Lothar Hoffmann que, ultrapassando a concepção wüsteriana de texto artificial, aproximou-o do texto elaborado na linguagem comum”. Assim ensina Hoffmann:

---

[8] Minha tradução: Linguagem especializada como meio de comunicação (1967). Da palavra especializada para o texto especializado (1988)

> O texto especializado é o instrumento ou o resultado de uma atividade comunicativa sócio-produtiva especializada. Compõe uma unidade estrutural e funcional (um todo) e está formado por um conjunto ordenado e finito de orações coerentes pragmática, sintática e semanticamente ou de unidades com valor de oração, que, como signos linguísticos complexos, de enunciados complexos do conhecimento humano e de circunstâncias complexas, correspondem à realidade objetiva (HOFFMANN, 1988, apud BRUMME, 1998, p.77. Minha tradução)[9].

Assim, entende-se o texto especializado como o texto configurado na comunicação verbal dialógica de um interlocutor para o outro em um contexto que versa sobre uma especialidade. Ainda que tal comunicação possa ser realizada de muitas maneiras e através de múltiplos canais, quando se fala em texto especializado, usualmente se considera o texto escrito, pois, como diz Lerat (1995, p.56) quem diz texto diz suporte (pedra, pergaminho, papel, meio eletrônico) e formalização da informação (frase, fórmula, figura, esquema, diagrama, organograma, etc.).

Talvez a noção exclusiva do texto escrito ainda seja uma herança da mentalidade vigente nos primórdios dos estudos terminológicos modernos, quando se pensava na linguagem da técnica e da ciência somente como uma realização escrita. De outro lado, a perenidade, a disponibilidade de acesso e, acima de tudo, as possibilidades de divulgação do texto especializado escrito asseguram sua precedência sobre o oral. Por isso, considerando que a comunicação e o registro do conhecimento são imprescindíveis para a

---

[9] *El 'text especializat' és l'instrument o el resultat de l'activitat comunicativa exercida en relació amb una activitat socioproductiva especialitzada. Forma una unitat estructural i funcional (un tot) i està format per un conjunt finit i ordenat d'oracions coherents pragmàticament, sintacticament i semànticament o d'unitats amb valor d'oració, que, com a signes lingüístics complexos d'enunciats complexos del coneixement humà i de circumstàncies complexes, corresponen a la realitat objectiva*.
HOFFMANN, L. (1988). Grundbegriffe der Fachsprachenlinguistik. In: *Der Ginko-Baum*. Germanistisches Jahrbuch für Nordeuropa, Siebente Folge. Helsinki.

constituição de um campo de saber, pode-se dizer que nenhuma área especializada sobrevive sem o texto escrito. No entanto, isso não significa que o texto oral seja excluído do conceito da linguagem especializada que é tanto falada como escrita de acordo com os respectivos recursos e exigências de cada modalidade de expressão.

## 3. O presente

O que entendemos por uma linguagem especializada hoje? Linguagem especializada e terminologia parecem conceitos muito próximos que muitas vezes se sobrepõem quando se fala em palavras técnicas, glossários, vocabulários e dicionários terminológicos ou quando se menciona a comunicação realizada em uma área do saber ou da experiência humana. O que não é de admirar, uma vez que um dos traços que parecem mais salientes nessa comunicação é a presença de uma terminologia desconhecida do leigo.

Quem se dedica a um ramo da ciência ou da técnica, exerce uma profissão, desenvolve atividades, seja na mais alta ou mais humilde esfera da sociedade, seja com fins lucrativos, assistenciais ou de simples lazer concorda que, na comunicação de seu conhecimento e de sua experiência, usa palavras específicas cujo sentido precisa ser explicado aos que não compartilham dos mesmos interesses. São palavras temáticas próprias da competência linguística dos que transitam em uma área determinada. Tais palavras pertencem à língua materna, elas se comportam morfológica e sintaticamente como qualquer outro item lexical do vocabulário do falante comum. No entanto, para saber utilizar essas palavras, é preciso recorrer à assistência de quem domina a área na qual elas são usadas, embora, para isso, não seja necessário ter acesso aos bancos escolares ou exercer uma ocupação sistemática e/ou remunerada.

A linguagem especializada é alvo de interesse tanto dos seus usuários diretos, os especialistas e os mediadores da informação, como dos linguistas e dos informatas. Para os especialistas, ela possibilita a configuração, expressão e comunicação do conhecimento. Para os mediadores da informação, entre esses, incluo professores, tradutores, redatores, comunicadores, ela se constitui na porta de

acesso da informação que será colocada à disposição do público interessado. Para os linguistas e informatas, sua análise, descrição e registro não se constituem em mero exercício acadêmico, mas visa a contribuir para a adequada transmissão do conhecimento e para o avanço da pesquisa do processamento da linguagem natural (PLN).

O desenvolvimento das linguagens especializadas é uma decorrência natural da especialização da atividade humana e da consequente formação de grupos de trabalho, fenômeno social comum desde os primórdios da história humana. Segundo Rastier (1995, p.47), na obra de Dante Alighieri, encontra-se a primeira menção de uma língua falada por profissionais do mesmo ofício: “depois da queda de Babel, apenas os arquitetos conseguiam se comunicar porque falavam uma língua que somente eles entendiam” (*De Vulgari Eloquentia,* I, 7). Nesse texto datado de 1305, associado ao conceito de especialização, já transparece a ideia, ainda hoje comum, de subsistema linguístico hermético de uma comunicação não entendida por alguém alheio ao círculo dos iniciados.

Modos de falar utilizados com o propósito de fazer segredo ou humor, fugindo das convenções linguageiras vigentes, surgem no âmbito de todas as camadas sociais e grupos de trabalho e as denominações de jargão e/ou gíria que lhes são atribuídas revelam conotação pejorativa. No entanto, é preciso enfatizar que não se trata de linguagens especializadas, mas de linguagens corrompidas, que constituem verdadeiras antilinguagens, como a linguagem do submundo das drogas e a linguagem das prisões. Além desses linguajares, também são comuns maneiras de falar elitistas, deliberadamente obscuras e incompreensíveis para o leigo como os assim chamados juridiquês, economês e jargão médico. A linguagem especializada não se enquadra nesse conceito, não é um jargão, nem uma gíria, mas também não é um subsistema linguístico reservado ao uso dos especialistas.

A denominação ‘subsistema’ atribuída à linguagem especializada foi bastante divulgada na segunda metade do século passado em publicações de renomados estudiosos da linguagem especializada, entre os quais Sager e Hoffmann. Contudo, no meu entendimento, o prefixo “sub” confere ao termo conotação negativa. O

conceito de sublingua é defendido por Zellig Harris (KITTREDGE; LEHBERGER, 1982, p. 01; p.231) que argumenta que em áreas de semânticas especializadas se formam, através de transformações da gramática da língua comum, subconjuntos fechados de construções que não pertencem à língua geral. Sob essa ótica o conceito de sublíngua se aproxima dos pressupostos adotados na Linguística Computacional e utilizados em metodologias que privilegiam as propriedades distribucionais em sistemas e aplicativos de PLN e de TA (Tradução Automática).

A noção de uma linguagem da ciência e da técnica separada da língua natural não é nova, foi amplamente difundida por ocasião do intenso intercâmbio científico e tecnológico posterior à Segunda Grande Guerra e consolidada pelos pioneiros da moderna terminologia e mesmo por aqueles que não partilhavam integralmente dos postulados wüsterianos. A lenta evolução desse conceito transparece em uma análise superficial do termo usado nos diferentes idiomas para a linguagem especializada. Os alemães foram os primeiros a enfatizar o seu aspecto profissional, *Fachsprache*, de *Fach* profissão, ofício e *Sprache* língua. Os ingleses se encarregaram de salientar-lhe a especificidade e as funções, difundindo-a internacionalmente como *special language* e *language for specific purposes – LSP*, língua especial ou língua para propósitos específicos. Enquanto os catalães e os espanhóis parecem marcar já na denominação a distinção entre língua e linguagem, *llenguatges d´especialitat, lenguajes de especialidad.*

Atualmente, é ponto pacífico, mesmo entre aqueles que continuam preferindo termo sublinguagem, que não se trata de uma língua diferente dos sistemas linguísticos naturais, mas sim de um uso peculiar da língua, isto é, uma linguagem, cuja configuração depende das múltiplas circunstâncias de sua realização. Outro aspecto da evolução do conceito é que tal linguagem não se limita à interação formal entre especialistas que usam um mesmo registro, mas é utilizada por interlocutores de diferentes hierarquias e graus de especialização em diferentes níveis de formalidade, desde o mais sofisticado cientista até o estudante e o leigo.

Em alguns meios linguísticos acadêmicos, ainda é discutida a ideia de que somente merecem a denominação de linguagens especializadas aquelas utilizadas por cientistas e seus pares, por profissionais e técnicos, isto é, somente aquelas que não dispensam um aprendizado formal para serem adquiridas. Essa mentalidade nos faz remontar à Academia Francesa do século XVII, que rejeitou de seu dicionário os termos referentes a artes, ciências e profissões porque não eram próprios de um homem de bem. Além disso, negando caráter de especialização a atividades de lazer e a ocupações exercidas por camadas sociais economicamente desfavorecidas, essa maneira de pensar revela uma tendência elitista estranha à mentalidade pluralista do mundo atual.

Ao falar no que se entende por linguagem especializada hoje, busco o apoio de Teresa Cabré, uma das vozes mais ouvidas no domínio da terminologia contemporânea. Cabré pondera que, dentro do sistema da língua geral, subsistem registros funcionais, empregados para realizar determinadas comunicações. Nesse sentido, o que nos habituamos a chamar linguagem especializada nada mais é do que um registro funcional utilizado em uma comunicação especializada. Observa a autora:

> [...] as chamadas linguagens de especialidade são registros funcionais caracterizados por uma temática específica, empregados em situações de características pragmáticas precisas, determinadas pelos interlocutores (basicamente o emissor), o tipo de situação em que são produzidas, e os propósitos ou intenções que se propõe a comunicação especializada. (CABRÉ, 1999, p.152. Minha tradução).[10]

---

[10] "[…]*, els anomenats llenguatges d'especialitat son registres funcionals caracteritzats per una temàtica específica, emprats en situacións de carcterístiques pragmàtiques precises, determinades pels interlocutors (bàsicamente l'emissor), el tipus de situació en què es produeixen, i els propòsits o intencions que es proposa la comunicació especialitzada.*" (CABRÉ, 1999, p.152)

Neste contexto, cabe lembrar que registro funcional não significa um subsistema linguístico separado, como Rondeau (1984, p.25-29) e alguns autores consideram, muito menos quer dizer jargão da ciência e da tecnologia. Registro funcional, de acordo com Halliday (1978, p.111), é a configuração de recursos semânticos que o membro de uma cultura associa com uma situação peculiar de comunicação. A esses recursos semânticos dever-se-ia acrescentar também os recursos sintáticos e morfológicos e pragmáticos.

A noção de registro funcional, aplicada à linguagem especializada é importante, porque põe em relevo tanto a situação real de uso como os interlocutores da comunicação especializada. Um registro funcional é uma variedade da língua comum associada a uma atividade específica, utilizada em circunstâncias dadas por falantes concretos em situações comunicativas dadas. Como manifestação da língua comum, a linguagem especializada partilha de todas as características do sistema linguístico que denominamos geral, seguindo o mesmo padrão e conformando-se aos mesmos parâmetros. Assim, ainda que dê preferência a alguns elementos léxicos e/ ou sintáticos de maneira particular, o uso especializado conta com o mesmo aparato linguístico da língua natural.

Consequentemente, o profissional e o cientista não se assemelham a programadores que utilizam uma sintaxe desenvolvida especialmente para a linguagem de máquina. Tampouco se expressam como matemáticos que usam sistemas peculiares de símbolos algébricos para apresentar o raciocínio e comprovar a solução dos problemas, também não podem ser incluídos entre os adeptos do formalismo lógico que usam sofisticadas combinações simbólicas. Ao contrário, os especialistas se comunicam com seus pares e com os leigos, utilizando a ferramenta linguística comum a todos os falantes, a língua materna. Suas preferências léxicas e gramaticais podem ser diferentes daquelas dos não especialistas, não obstante, essas mesmas opções linguísticas poderiam, teoricamente, ser realizadas por qualquer outro usuário da língua em qualquer contexto de comunicação.

Isso não significa que não haja peculiaridades decorrentes dos propósitos de comunicação, uma vez que não se pode deixar de salientar a relevância dos traços que compõem a especificidade e

marcam a especialização. Em primeiro lugar, destacam-se fatores primários inerentes à natureza do campo de conhecimento e da atividade profissional que perfazem o contexto temático de uma área dada. Tal contexto é articulado pelo estatuto do destinador e o do destinatário, a adequação dos propósitos da comunicação e os aspectos circunstanciais do evento comunicativo. Em resumo, a especialização se manifesta na expressão linguística, tomada em sentido lato, levada a efeito em um contexto sócio-cultural complexo, no qual aspectos temáticos e pragmáticos se entrelaçam.

A especialização resulta de um complexo contexto, no qual não só o aspecto do conteúdo é importante, isto é, não é o significado léxico o único e exclusivo fator que lhe confere o caráter específico. A referência a um conjunto de objetos empíricos não é suficiente para caracterizar uma área temática, posto que um domínio ou campo de conhecimento se compõe de um amálgama de elementos que conformam o universo do discurso do especialista. Nesse complexo de elementos, destaca-se, junto ao aspecto semântico, expresso na terminologia, o componente cognitivo, isto é, a perspectiva em que o segmento do universo enfocado pelo especialista é considerado.

Dessa maneira, na matemática e na filosofia, a realidade é contemplada sob a ótica formal da conceptualização. Nas ciências naturais, a necessidade de categorização dos componentes da natureza se traduz no caráter taxonômico predominante. No direito, área em que a sociedade é contemplada com fim de equilibrar, organizar e regulamentar as relações humanas, a orientação prescritiva é privilegiada. Ao passo que a informação sobre a maneira de utilizar os bens do universo se constitui na preocupação principal das áreas técnicas, artísticas, artesanais e ocupacionais, como eletrônica, música, tecelagem, segurança pública, disposição do lixo ou coleta de resíduos recicláveis. De acordo com o enfoque próprio da área, o especialista examina o universo e o codifica, modulando sua linguagem conforme as condições circunstanciais do evento comunicacional em curso.

Nesse contexto, o assunto tratado pode ser mais ou menos complexo, pode ser abordado em diferentes graus de profundidade, em maior ou menor nível de especialização. O meio de comunicação pode ser a língua oral ou a escrita, passando pelo simples boletim de

ocorrência policial à revista acadêmica, do manual ao livro texto, da comunicação em encontro restrito à conferência de seminário de alcance internacional.

Enfim, em um leque ilimitado de realizações, a linguagem especializada reflete os propósitos da comunicação que, por sua vez, atende aos objetivos estabelecidos pela área temática em questão. À função denominativa básica da linguagem especializada se somam outras funções tais como a informativa, a argumentativa, a descritiva e a prescritiva. Nenhuma dessas funções tem exclusividade, porém, para os propósitos de cada área, uma função exerce predominância sobre as outras, conforme as condições efetivas do processo comunicacional.

Em suma, o caráter especializado da linguagem não é conferido pelo tema enfocado, nem pela terminologia empregada, mas antes de tudo pela maneira de tratá-lo na comunicação. Como observa Cabré (1999, p.154), um conteúdo tradicionalmente científico, apresentado de maneira banal, sem relação com uma estrutura sistemática preestabelecida de significação, não se constitui em matéria de comunicação especializada. De maneira inversa, quando descrito de maneira precisa, com referência a um esquema científico determinado ou a uma estrutura normativa estabelecida, o conteúdo torna-se objeto de conhecimento especializado e, portanto, de comunicação em linguagem especializada. Por isso, Alain Rey, o renomado linguista, lexicógrafo dos dicionários Robert e autor de uma basilar tese sobre terminologia (1992), diz que [11]

> O que torna "especial" uma língua de especialidade não é somente seu vocabulário (no discurso), seu léxico (no sistema), sua terminologia (no plano conceitual e cognitivo), mas é antes de tudo enquanto representante linguageiro de uma coerência conceitual. (REY, 1991, p. ix. Minha tradução.)

[11] *Ce qui rend "spéciale" une langue de spécialité, ce n'est pas seulement son vocabulaire (em discours),son lexique (en système), sa terminologie (sur le plan conceptuel et cognitif), mais c'est avant tout cela, en tant que représentant langagier d'une coherence conceptuelle.(REY,* 1991, p. ix)

De acordo com Guiomar E. Ciapuscio (1998, p.43), linguista argentina cuja autoridade não pode ser contestada, compreendo um "texto como uma unidade linguístico-comunicativa, resultado de uma conceptualização por parte de um falante e proposta, ao mesmo tempo, à interpretação de eventuais destinatários"*.* Assim, ainda que apresente características que revelam à primeira vista sua especificidade, o texto especializado, como manifestação linguística, não difere essencialmente do texto não especializado.

Com efeito, como expressão da linguagem de especialidade, um texto revela o padrão sintático-pragmático de manifestações linguístico-textuais na língua geral, proposto por vários autores que se debruçaram sobre linguística do texto, entre os quais destaco Beaugrande e Dressler (1981, p.3-11). Tal padrão pode ser sintetizado num conjunto mínimo de condições que configuram a comunicação como um texto. São condições básicas que abrangem aspectos referentes à estrutura textual, às qualidades de conteúdo e à atitude dos interlocutores.

Assim, pela coesão, os elementos lexicais que compõem o texto se interligam mutuamente, compondo uma sequência lógica de acordo com as convenções gramaticais. A coerência assegura que o conteúdo expresso nos conceitos e relações da tessitura superficial possa ser recuperado a partir do universo cognitivo ativado pelo texto.

À coesão e coerência, acrescentam-se a informatividade, a situacionalidade e a intertextualidade como características que equacionam a informação apresentada com a expectativa do destinatário, no sentido da novidade e recentidade do que é informado e, também, quanto à sua relevância para a situação e ainda no que diz respeito à sua ligação com outros textos conhecidos. Além disso, é importante que o destinador tenha a intenção de produzir o texto e que o destinatário o receba como tal. A intencionalidade e a aceitabilidade são condições que dizem respeito às atitudes dos interlocutores. Concebido dentro desses padrões, o texto especializado, em suas propriedades essenciais, não difere de um texto não especializado.

A especialização do texto se origina da combinação de dois critérios: critério temático e o critério pragmático, que se conjugam de acordo com a função comunicativa própria da área e especificamente

visada pelo autor. Mesmo que a função referencial cognitiva seja predominante, a função comunicativa do texto especializado não é una, mas múltipla, várias outras funções se manifestam e se entrelaçam. Em áreas essencialmente científicas e técnicas, a função majoritária é apreender os fenômenos naturais para produzir e divulgar conhecimentos. Ao passo que, em outros domínios de natureza humana e social, a função prescritiva se salienta na emissão de ordens e instruções.

Dessa maneira, se constata que "há uma estreita relação entre as funções escolhidas e o âmbito temático correspondente" como bem observam Möhn e Pelka (1984, p.5)[12], citados por Hoffmann (*apud* BRUMME, 1998, p.108). Tal relacionamento estreito repercute na organização textual, na seleção sintática e no uso da terminologia. Essas preferências igualmente se condicionam às circunstâncias da interação verbal dos atores do evento comunicativo, dependendo, portanto, das peculiaridades do destinador e do destinatário.

O critério temático diz respeito, primordialmente, ao tema da área enfocada, abrangendo não só aspectos semânticos do assunto como as características cognitivas da área. A dimensão pragmática, relativa aos participantes da comunicação, acrescida das várias e complexas circunstâncias que compõem o evento, contribui para imprimir traços específicos no texto. Assim, um texto sobre o DNA, com vistas aos procedimentos da ação de investigação da paternidade, é concebido na perspectiva das implicações legais do tópico e redigido conforme os padrões ditados pela área jurídica. Enquanto um texto sobre o DNA, como o composto químico ácido desoxirribonucléico e seu papel na transmissão dos caracteres hereditários é um texto de genética, e apresenta a composição textual própria da área científica. Esse texto somente será considerado jurídico, se for produzido com o enfoque no papel do teste de DNA nos processos judiciais, salientando a implicação do elemento genético no procedimento legal.

Preocupados com o conceito de especialização, muitos autores têm investigado o grau de especificidade do texto, salientando

---

[12] MÖHN, D.; PELKA, R. (1984) *Fachsprachen. Eine Einführung*. Tübingen. (*apud* HOFFMANN, 1998)

componentes morfossintáticos ou aferindo a gradação da quantidade de informação que o texto contém. No entanto, nem uns, nem outros examinam os elementos que compõem a natureza da especificidade. Esses pesquisadores buscam avaliar a extensão e a essência do conhecimento especializado posto à disposição do usuário. Em tal análise, não visam aos elementos distintivos que caracterizam o texto como expressão do conhecimento de uma área particular, mas aos elementos que tornam a expressão do conhecimento mais acessível ao público alvo.

Esses autores, em sua maioria inspirados em Hoffmann (*apud* BRUMME, 1998, p.56-69), admitem dois eixos de classificação na avaliação do grau de especificidade de um texto: um horizontal, baseado no critério temático, e outro vertical, baseado no grau de abstração da linguagem. Desse modo, no primeiro eixo alinham-se as mais variadas disciplinas, atividades e profissões, enquanto, na escala vertical se consideram cinco níveis de abstração e de realizações discursivas. No patamar mais alto, está a linguagem científica de grande sofisticação, usada em reuniões de especialistas, como em congressos de Física Quântica, ao passo que no nível inferior, encontra-se a interação verbal da rotina de operários especializados em seu ambiente de trabalho, como uma fábrica de motores ou uma plataforma de petróleo.

Nesse contexto, a clássica análise de Hoffmann (ibidem, p.62-69) procura equacionar o grau de especialização e o grau de abstração, que é aferido pelo uso frequente da terminologia e pela preferência das construções sintáticas utilizadas. De acordo com tais critérios, Hoffmann classifica as linguagens de especialidade em cinco categorias principais, que ele mesmo afirma não serem plenamente satisfatórias: linguagem das ciências fundamentais, das ciências experimentais, das ciências aplicadas e da técnica, da produção material e do consumo.

Ciapuscio (1998, p.43-65) abre novas perspectivas para a questão teórico- metodológica dos graus de especialização do texto. Nesse sentido, propõe outro ângulo de análise, a variação conceitual, e usa, para representar a informação conceitual e formal do signo linguístico, o ‘esquema’ como instrumento. Dentro da proposta de redes conceituais, o

‘esquema’ permite avaliar como conceitos e relações são, de um lado, inscritos no texto e de outro, apreendidos pelo usuário à medida que ele segue o raciocínio que o autor desenvolve ao longo do texto.

Dessa forma, Ciapuscio avalia o grau de abstração conceitual em diferentes categorias de textos que versam sobre a mesma temática, mas que se dirigem aos mais diferentes perfis de usuários. Para tanto, a autora mede a profundidade e abrangência do conhecimento veiculado através de dois parâmetros. O primeiro é revelado pelo uso da terminologia específica, enquanto o segundo corresponde à variação terminológica realizada na forma de sinônimos, paráfrases e explanações.

A autora examina como a variação conceitual do termo se adapta à variação discursiva, a fim de modular o grau de densidade da informação a ser oferecida ao usuário, de maneira que o texto se torne mais ou menos transparente. A análise mostra que, quando não há variação da terminologia, o texto revela um maior grau de densidade do conhecimento especializado proposto ao usuário.

Desse modo, o contexto da situação do destinatário é o fator de maior peso na avaliação do grau de especialização do texto, influindo mais na determinação do grau de abstração do que a temática. Sob esse prisma, a seleção diferenciada de termos e a variação denominativa caracterizam os textos de divulgação e revelam a imagem que o autor faz das condições cognitivas de seu público alvo.

Cabré (1993, p.151-156; 1999, p.156-179) distingue os textos de língua comum e os textos de especialidade, apontando elementos linguísticos e pragmáticos. Do ponto de vista linguístico, a autora considera as características lexicais, morfológicas e sintáticas. Sob o ângulo da pragmática, analisa produtores e destinatários, situações comunicativas e propósitos visados. Quanto à funcionalidade, Cabré enfoca as funções prioritárias às quais os textos se propõem e destaca as intenções pragmáticas do autor.

A autora pondera que o texto especializado mostra preferências que variam segundo as diferentes áreas de especialização, desde a seleção do vocabulário e padrões morfossintáticos, até estruturas textuais e discursivas. Todavia, ela também adverte que não há justificativa para falar da existência de um código linguístico diferente

daquele usado nos textos não especializados. Porém, mostra que os aspectos pragmáticos permitem distinguir um tipo de texto do outro, singularizando o texto especializado no que diz respeito à temática, aos usuários e ao contexto de comunicação.

Em resumo, sob a luz do que foi aqui exposto, a realização linguística do texto especializado, nela compreendida sua estruturação gramatical, textual e terminológica e ainda sua formatação gráfica, depende de fatores temáticos e pragmáticos. A influência desses fatores se faz sentir tanto na ativação do valor especializado das palavras que, no evento comunicativo, desempenham a função de vetor da transmissão da informação, da instrução, do mandamento, da sugestão e do conselho, como na seleção dos elementos lexicais que os articulam na estrutura sintática e na configuração discursiva. Tal seleção demonstra a preferência por itens lexicais que rigorosamente não têm nenhum traço de especialização ou por determinadas construções sintáticas, caracterizando a linguagem especializada de uma área dada e configurando o pano de fundo para a articulação das palavras que exprimem o conteúdo do texto, os termos.

## 4. O futuro

Ao discorrer sobre o que me parece consensual hoje no entendimento do conceito de linguagem especializada e do relacionamento de sua pesquisa com a pesquisa terminológica na época em que vivemos, recorri a linguistas deste e do outro lado do Atlântico que há longos anos se debruçam sobre a área e que muito têm contribuído para seu avanço. O fato de não ter mencionado outros autores não significa menosprezo pelo trabalho dos mais jovens, na pesquisa e/ou na idade, pelo contrário, eles trazem novas luzes e apontam novos caminhos que enriquecem e fertilizam a área da linguística e da terminologia. O tema aqui abordado pode alimentar, como aliás já alimenta, inúmeras teses, seminários e encontros, e merece que se formem grupos de trabalho para aprofundar e ampliar os pressupostos teóricos que poderão orientar a configuração de um estudo científico das linguagens especializadas, recorrendo à linguística geral e à linguística aplicada sem ignorar a terminologia.

Em tal contexto, será fortalecido o entendimento de que o texto especializado é um todo complexo no qual não se pode abstrair o conteúdo temático da formulação verbal e será renovada a crença de que um sem o outro não poderá subsistir, uma vez que se forem fragmentados perderão seu caráter específico. Assim, se de um lado, não existem textos especializados sem termos, por outro lado, uma terminologia fora da linguagem especializada não tem razão de ser. O estudo científico das linguagens especializadas, isto é, a linguística das linguagens especializadas e o estudo científico dos termos, isto é, a terminologia, não são áreas separadas e, sim, intimamente relacionadas e precisam ser integradas. Por essa razão, repito com Sager (2007, p.205) que a separação do estudo das línguas especializadas e a terminologia não se justifica.

À guisa de conclusão, do artigo fundamental artigo *Textes et Termes* da revista *Meta*, transcrevo a citação que reúne o pensamento dos dois grandes entusiastas do estudo científico das linguagens de especialidade, Kocourek e Hoffman[13]:

> A tônica sobre o lado textual da língua erudita foi geralmente caracterizada por um desinteresse do lado terminológico da língua erudita, cuja importância era considerada ultrapassada. Essa tendência é compreensível na etapa do desenvolvimento que seguiu imediatamente à dominação quase absoluta da terminologia nos estudos terminológicos. Mas, numa ótica equilibrada, a marginalização da terminologia não tem fundamento. (KOCOUREK, 1990a, ch.1.3.4, cf.,HOFFMANN, 1988: 120. Minha tradução)

---

[13] “L’accent sur le côté textuel de la langue savante était généralement caractérisé par un désintérêt du cote terminologique de la langue savante, dont l’importance était considerée comme dépassée. Cette tendance est compréhensibloe à l’étape du dévelopment qui suit immédiatement la dominance presque absolue de la terminologie dans les études terminolinguistiques. Mais dans une optique équilibrée, la marginalisationde la terminologie est sans fondement KOCOUREK, 1990a, ch.1.3.4, cf.,HOFFMANN, 1988: 120).

## Referências

AUSTIN, J. L. (1978) *How to do things with words*. Cambridge, Mass.: Harvard University Press. 3[a] ed. 169 p.

BEAUGRANDE, R. de; DRESSLER, W. (1981) *Introduction to text linguistics*. English translation by Robert Beaugrande. New York: Longman. 270 p.

BENVENISTE, E. (1989) *Problemas de linguística geral II*. Campinas: Pontes. Trad. Eduardo Guimarães *et alii.* 294p.

BRUMME, J. (ed) (1998) *Lothar Hoffmann*. LLengatges d'especialitat. Selecióde textos Barcelona: IULA/UPF. 284 p.

CABRÉ, M.T. (1993) *La terminología*: Teoria, metodología, aplicaciones. Barcelona: Antártida/Empúries. 529 p.

CABRÉ, M.T. (1999) *La terminología:* representación y comunicación: elementos para una teoria de base comunicativa e otros artículos. Barcelona, IULA, 1999. 369p.

CIAPUSCIO, G. E. (1998) La terminología desde el punto de vista textual: selecció, tratamiento y variación. *Organon*, v.12, n.26, 1998. p.43-65.

FINATTO, M. J. B. (2001) *Definição terminológica*: fundamentos teóricos e metodológicos para sua descrição e explicação. Porto Alegre, UFRGS. 395p. (Tese, Doutorado em Letras, Estudos da Linguagem: Teorias do Texto e do Discurso).

HALLIDAY, M. A. K. (1978) *Language as social semiotic*. London: Edward Arnold, 75p.

HOFFMAN, L., KALVERKÄMPER, WIEGAND, H. E.,(ed.) (1999) : *Fachsprachen. Languages for Special Purposes,* Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft An International Handbook of Special Language and Terminology Research. Berlin: Walter de Gruyter, coll. «Handbücher zur Sprach- und Kommunikationswissenschaft », vol. 14-1, t. 1, 1369 p.

HOFFMANN, L (1987). *Kommunikationsmittel Fachsprache:* eine Einführung. Berlin: Akademie. 307p.

HOFFMANN, L. (1988) *Vom Fachwort zum Fachtext*. Tübingen: Gunter Nar. 265 p.

KITTREDGE, R.; LEHBERGER, J. (1982) *Sublanguage*: studies of language in restricted semantic domains. Berlin: Walter de Gruyter. 240p.

KOCOUREK, R. (1991a) *La Langue Française de de la Technique et de la Science:* vers une linguistique de la langue savante. Wiesbaden: Brandstetter. 327 p.

KOCOUREK, R. (1991b) Textes e Termes, *Meta*, v.36, n.1, 1991, p.71-76.

LERAT, P.(1995) *Les Langues spécialisées.* Paris: PUF. 201 p.

RASTIER, François. (1995) Le terme: entre ontologie et linguistique. *La Banque des Mots*. n.7, 1995, numéro spécial. Premières Rencontres Terminologie et Intelligence Artificielle. p.35-65.

REY, Alain. (1991) Avant-Propos. In: KOCOUREK, R. (1991) *La Langue Française de de la Technique et de la Science: vers une linguistique de la langue savante.* Wiesbaden: Brandstetter. p. vii-xiii.

REY, Alain. (1992) *La terminologie:* noms et notions. Paris: P.U.F. Universitaires de France. 127p. (Que sais-je?)

RONDEAU, Guy. (1984) *Introduction à la terminologie.* Québec: Gaetan Morin. 227p.

SAGER J. C. (2007) La evolución de los lenguajes de especialidad y la terminología. In: LORENTE, M.; ESTOPÁ, R.; MARTÍ, J.; TEBÉ, C. *Estudis de linguística i de linguística aplicada en honor de María Teresa Cabré Castellví*. Vol. I-De mestres. Barcelona: IULA/UPF, p. 105-122.

SAGER, J.C.; DUNGWORTH, D.; P. F. MACDONALD (1980) *English Special Languages: Principles and practice in science and technology.* Wiesbaden: Brandstetter, 368 p.

WÜSTER, Eugen. (1974) Die allgemeine Terminologielehre - ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenshaften. *Linguistics*, n.119. La Hague: Mouton. p.61-106.

# CHUTANDO O BALDE OU BATENDO AS BOTAS? PROCESSAMENTO DE LINGUAGEM NATURAL E EXPRESSÕES MULTIPALAVRA NA LINGUAGEM COTIDIANA E CIENTÍFICA

*Aline Villavicencio*[1]

*Carlos Ramisch*[2]

***Resumo**: Expressões Multipalavra são parte integrante das linguagens humanas. Apesar de sua importância ter sido reconhecida, suas características heterogêneas têm-se revelado um desafio para análises linguísticas e computacionais. Neste capítulo, examinam-se as Expressões Multipalavras e alguns dos desafios que elas representam para essas áreas. A seguir, apresenta-se uma variedade de abordagens para detectar e lidar com essas expressões de forma semi-automática, trazendo exemplos de seu tratamento em corpora de linguagens especializadas.*

## 1. Introdução

Durante uma interação em língua natural, locutores nativos utilizam majoritariamente construções convencionais adotadas pelas suas comunidades linguísticas para expressar com precisão ideias e conceitos complexos. A expressão *bode expiatório*, por exemplo, exprime de maneira compacta o conceito de "pessoa escolhida arbitrariamente para levar sozinha a culpa de uma calamidade, crime ou evento negativo (que geralmente não cometeu)[3]". Frequentemente tais conceitos, como o de bode expiatório, não podem ser compactados em uma única palavra e, consequentemente, Expressões Multipalavras (EMs) são necessárias para transmiti-los de forma eficiente e concisa.

---

[1]Instituto de Informática, Universidade Federal do Rio Grande do Sul, avillavicencio@inf.ufrgs.br

[2]Instituto de Informática, Universidade Federal do Rio Grande do Sul, ceramisch@inf.ufrgs.br

[3]Fonte: pt.wikipedia.org/wiki/Bode_expiatório

EMs têm recebido a atenção de pesquisadores em várias disciplinas, como em Linguística (Fillmore et al., 1988; Fillmore, 2003; Nunberg et al., 1994; Pawley, 2007), em Psicolinguística (Cacciari, 2003; Jackendoff, 1997; Wray, 1999, 2002), e em Processamento de Linguagem Natural (Calzolari et al. 2002; Katz e Giesbrecht, 2006; McCarthy et al., 2003; Sag et al., 2002). Por serem de natureza diversa, EMs formam um conjunto heterogêneo de fenômenos linguísticos que inclui desde colocações a expressões idiomáticas, conforme exemplificado na Tabela 1. Como consequência, o processamento uniforme das EMs, ou seja, as tentativas de caracterizá-las completamente e de definir precisamente as suas fronteiras, constitui um grande desafio para as disciplinas que as estudam, na medida em que EMs podem ser vistas sob tantos ângulos diferentes quanto as várias construções linguísticas que o conceito engloba. Para Moon (1998), "não há um fenômeno unificado que se possa descrever, mas sim um complexo de atributos que interagem de formas diversas, muitas vezes desordenadas, e que representam um amplo contínuo entre o não composicional (ou idiomático) e grupos composicionais de palavras". Assim, autores diferem, entre si, nas definições de EM, dependendo dos aspectos particulares que estão sendo enfatizados e dos grupos de palavras e construções considerados.

Nesse contexto, Sag *et al.* (2002) definem EMs como "interpretações idiossincráticas que cruzam os limites (ou espaços) das palavras". Eles se concentram na diferença entre a interpretação de uma EM (ex. *bode expiatório*) como um todo e os significados isolados das palavras individuais que compõem a EM (*bode* e *expiatório*), e consideram como EMs um grande número de fenômenos, tais como expressões fixas, compostos nominais e construções verbo-partícula. Ainda nesta linha, Calzolari *et al.* (2002) definem uma EM como "uma sequência de palavras que atua como uma única unidade, em algum nível de análise linguística", e que possui algumas das seguintes características:

1. transparência sintática e/ou semântica reduzida;
2. composicionalidade reduzida;
3. flexibilidade sintática reduzida;
4. violação de regras sintáticas gerais;
5. elevado grau de lexicalização;
6. elevado grau de convencionalidade.

Tabela 1 – Exemplos de alguns tipos de Expressão Multipalavra.

| **Tipo de EM** | **Exemplo** |
| --- | --- |
| Colocação | *sal e pimenta, bagagem emocional, strong tea* |
| Expressão idiomática | *cutucar a onça com vara curta, rock the boat* |
| Nome próprio | *São Paulo, Estação da Sé, Manchester United* |
| Composto nominal | Subst + *de* + subst: *pitada de sal, carro de polícia, faca de pão, flock of geese*<br>Subst + adj: *taxas federais, momento cívico, friendly fire* |
| Substantivo + preposição | *cópia de, viagem para* |
| Verbo de suporte (VS) | VS com sujeito: *fire burns*<br>VS com objeto: *tomar uma decisão*, *take a shower*<br>VS ditransitivo: *dar conselho a* [SN]<br>VS com adjetivo: *ficar biruta, get naked* |
| Preposição de suporte | *em perigo, at risk* |
| Verbo preposicionado (VP) | VP intransitivo: *procurar por* [SN], *look into* [NP]<br>VP transitivo: *confiar* [SN] *a* [SN] |
| Construção verbo-partícula (CVP) | CVP intransitiva: *sit down, shut up*<br>CVP transitiva: *take* [NP] *out, take* [NP] *off* |
| CVP com preposição | *put up with* |
| Verbo + subst + prep | *take advantage of* |
| Adjetivo + preposição | *similar a, fond of* |
| Sequência lexical | *Eu gostaria de, I don't know whether* |
| Expressão fixa | *ad infinitum, in vitro* |
| SP sem artigo | *de cama, em reunião, at school* |

Wray considera como EM "qualquer tipo de unidade linguística que tenha sido considerado *formulaica* em qualquer campo de pesquisa" (1999, p.9), enquanto Fillmore (2003) se concentra na noção de um Falante/Ouvinte Inocente (*Innocent Speaker/Hearer* - ISH) com o conhecimento apenas sobre palavras individuais e relações palavra-a-

palavra. Nesse contexto, EMs englobam “qualquer expressão linguística envolvendo mais de uma palavra que exija que um intérprete - humano ou computacional - tenha mais do que as habilidades de um ISH”. Fillmore, Kay e Connor (1988) observam que esta ideia introduz uma distinção entre o que um falante de uma língua é capaz de deduzir automaticamente a partir da linguagem e que ele deve conhecer previamente ou armazenar explicitamente. Idealmente, o número de informações as serem previamente sabidas deve ser minimizado, de forma que a maioria dos conhecimentos necessários ao uso da língua seja deduzível. Fillmore considera extremista o seu conceito de EM, pois o mesmo inclui uma ampla gama de construções gramaticais e de configurações de palavras listáveis, até mesmo sequências como *cópia de* em *Recebi uma cópia de alguns dos memorandos*. No entanto, o autor observa que Jackendoff (1997) vai ainda mais longe e incorpora também à definição de EM poemas memorizados e frases familiares de comerciais de TV. Outros autores utilizam a noção de frequência e definem EMs como sequências ou grupos de palavras que co-ocorrem com mais frequência do que seria esperado se essas palavras fossem totalmente composicionais, e que podem, portanto, ultrapassar fronteiras sintagmáticas (Dias et al., 1999; Gil e Dias, 2003; Tremblay et al.; 2007). Isso incluiria também fórmulas de saudação como “*Tudo bem? Como vai você?*” e sequências lexicais, como *eu não sei se*.

Portanto, decidir se uma sequência de palavras forma uma EM ou não pode ser uma tarefa difícil que envolve não somente a avaliação de algumas das propriedades gerais de EMs como também a avaliação de características de tipos específicos. Por exemplo, Nunberg, Sag e Wasow (1994) descrevem as seguintes características para expressões idiomáticas, das quais somente a primeira é obrigatória:

1. O seu significado é convencionalizado uma vez que não pode ser totalmente inferido a partir dos significados das palavras componentes quando usadas separadamente umas das outras.
2. Elas podem ser sintaticamente mais rígidas e menos flexíveis.
3. Elas podem envolver figuras de linguagem como metáforas ou hipérboles.

4. Elas podem ser usadas para descrever proverbialmente e informalmente uma situação recorrente, referenciando situações ou objetos familiares de uma forma mais coloquial.
5. Elas podem expressar um estado afetivo ou um julgamento sobre algum tópico.

No contexto de textos de um domínio específico, tem-se uma definição importante relacionada a EMs, a de *termo*. Segundo Krieger e Finatto (2004), termos são unidades lexicais que expressam conceitos abstratos do domínio. Existem várias diferenças entre EMs e termos: termos podem ser compostos por uma única palavra ou por múltiplas, como locuções nominais e verbais, enquanto EMs são inerentemente compostas por duas ou mais palavras. Em segundo lugar, EMs estão presentes tanto na linguagem técnica e científica quanto na linguagem cotidiana de propósito geral, enquanto termos são, em princípio, particulares às linguagens especializadas. Em terceiro lugar, os métodos computacionais disponíveis para lidar com EMs em textos genéricos nem sempre podem ser aplicados para lidar com *corpora* especializados e para extração de terminologia. Por outro lado, EMs e termos têm também alguns aspectos comuns na medida em que ambos possuem uma semântica não convencional e que ambos constituem um desafio para os sistemas de Processamento da Linguagem Natural (PLN).

Neste capítulo, apresenta-se uma visão geral de EMs, analisando suas dimensões linguísticas, e os desafios que elas representam para seres humanos e para sistemas computacionais (seção 2). Em seguida, descreve-se o estado atual da pesquisa sobre EMs da perspectiva do PLN (Seção 3), e em particular as diferentes abordagens para identificar e modelar EMs em *corpora* especializados (Seção 4).

## 2. EMs sob o microscópio

EMs em geral possuem características idiossincráticas (Baldwin, 2006; Korhonen, 2006; McCarthy, 2006), seja em nível lexical, sintático, semântico, pragmático ou estatístico e, possivelmente, em mais de um nível, como mostra a Tabela 2 (adaptada de Baldwin, 2006). Nas

próximas seções, cada um desses níveis de interpretação será analisado em detalhe, de forma a caracterizar a flexibilidade e variabilidade das EMs a cada um desses níveis.

Tabela 2 – Exemplo de idiossincrasias de diferentes níveis para algumas EMs.

| **EM** | **Idiossincrasia** | | | | |
|---|---|---|---|---|---|
| | **Léxica** | **Sintática** | **Semântica** | **Pragmática** | **Estatística** |
| ad infinitum | √ | ? | ? | ? | √ |
| em casa | | √ | | | |
| primeiros socorros | | | √ | | ? |
| chá forte | | | | | √ |
| feliz aniversário | | | | √ | √ |
| cat's cradle | √ | √ | √ | | ? |

## *2.1. Caracterização lexical*

Algumas EMs, como *ad infinitum*, contêm palavras componentes que não podem ser utilizadas isoladamente (Baldwin, 2006; Sag *et al.*, 2002). Em língua portuguesa, existem diversas expressões lexicalmente idiossincráticas que, a exemplo da anterior, são emprestadas de do latim, como *ad hoc*, *et al.*, *in vitro*, *stricto sensu*. O fato de que a expressão é na verdade um empréstimo do latim pode explicar a razão pela qual as palavras componentes não são empregadas isoladamente na linguagem cotidiana. Todavia, existem também EMs desse tipo para as quais uma das palavras que a compõe não pode ser usada separadamente, apesar de possuir uma semântica bem definida em português, como as locuções conjuntivas *apesar de, no entanto, não obstante* (as palavras *apesar, entanto* e *obstante* não são encontradas separadamente).

As EMs lexicalmente fixas geralmente podem ser atribuídas, como um todo, a uma classe morfossintática precisa (por exemplo, conjunção), e costumam ser processadas pelos sistemas de PLN como um lexema (ou seja, uma entrada de dicionário) que contém espaços. Essa é certamente a maneira mais eficiente de tratar essas construções, pois qualquer abordagem composicional implicaria num tratamento independente para cada uma das partes da EM. Um sistema de análise sintática, por exemplo, seria induzido a cometer erros do tipo superanálise (*overgeneration*), ou seja, palavras como *ad* e *vitro* seriam corretamente reconhecidas mesmo se usadas separadamente. Contudo, esse tipo de abordagem, denominado *palavras-com-espaços*, não se aplica às EMs lexicalmente flexíveis, ou seja, àquelas cujas idiossincrasias se manifestam em níveis superiores (sintático, semântico, etc.).

## *2.2. Caracterização sintática*

As EMs sintaticamente idiossincráticas não seguem as regras usuais da gramática de uma língua, por exemplo, na expressão *em reunião,* espera-se encontrar um artigo precedendo o substantivo no sintagma nominal. Os componentes dessas EMs podem apresentar entre si relações complexas, e vão desde manifestações produtivas com grande flexibilidade e variabilidade na sua forma até manifestações que não possuem nenhuma variação possível, como *bater as botas* (por exemplo, *bateu/bateram/bate as botas*,...) e *by and large* (*by and *short/*largest*), respectivamente. Apesar de permitirem certo grau de flexibilidade sintática, as expressões dessa classe não permitem que se apliquem todas as modificações passíveis de serem aplicadas a expressões totalmente composicionais. No caso de *bater as botas*, por exemplo, pode-se conjugar o verbo *bater* de acordo com o contexto de uso sem que, entretanto, seja possível transformar a expressão para a voz passiva (**as botas foram batidas*).

Além disso, algumas EMs possuem componentes variáveis, mas que devem obedecer a certas restrições, como na expressão *keep [SN] on [SN]'s toes* onde tanto o SN objeto quanto o pronome possessivo podem variar, mas devem concordar (por exemplo, *She*

*kept us on our/*his/*their toes during the whole interview*). No entanto, esse grau de flexibilidade parece não ser previsível a partir da estrutura sintática em si, como no caso das expressões idiomáticas verbais *chutar o balde* e *bater as botas*. Ambas são formadas por verbos transitivos com objetos diretos, porém diferem no grau de flexibilidade sintática: enquanto a primeira aceita variação na forma (*chutou mesmo/de vez o balde*), a segunda tem forma rígida. Além disso, o resultado da combinação sintática de componentes nem sempre é previsível a partir do comportamento dos componentes, como se pode perceber em *wine and dine*, que é uma expressão transitiva (por exemplo, *He likes to wine and dine his guests*) que resulta da coordenação de dois verbos intransitivos (*wine* e *dine*), como apontado por Baldwin (2006). Por outro lado, mesmo nos casos em que as características sintáticas são facilmente estabelecidas, o problema pode estar em decidir qual a categoria morfossintática que os constituintes devem ter, como é o caso da EM *ad hoc*.

Mas será que o comportamento sintático das EMs é realmente tão imprevisível? Em uma avaliação da distribuição de flexibilidade de EMs, considerando uma amostra aleatória das 100 expressões mais frequentes em inglês, Villavicencio e Copestake (2002) constataram que os 43 casos de expressões verbais se dividiam em 22 grupos, cada um definindo um padrão de variação distinto. No entanto, a distribuições das instâncias nos grupos não é uniforme: os maiores grupos eram aqueles com características regulares, como o grupo de verbos transitivos com complementos definidos (*chutar o balde*). Quanto mais variabilidade sintática, menor o número de elementos do grupo, por exemplo, o grupo com verbos variáveis e elementos opcionais (*touch/find/strike a [raw] nerve*). Além disso, para os casos mais flexíveis, mesmo que uma forma canônica da EM possa ser encontrada, frequentemente ela também ocorre nas formas não canônicas (por exemplo, na voz passiva). Essas variações, segundo Riehemann (2001), são responsáveis por 25% das ocorrências de um determinado tipo de expressão em *corpora*.

## *2.3. Caracterização semântica*

O significado de uma EM varia desde as mais transparentes semanticamente (por exemplo, *bom dia*) até casos mais idiomáticos (por exemplo, *João sem braço,* que se refere a alguém preguiçoso, malandro ou dissimulado). A distinção entre EMs semanticamente opacas e transparentes pode ser formalizada através da noção de composicionalidade, que se refere à capacidade de inferir a interpretação de uma sequência de palavras a partir da semântica de cada uma das palavras que a compõe. O grau de composicionalidade de uma EM pode ser difícil de aferir: entre as expressões idiomáticas e aquelas que podem ser completamente modeladas através da polissemia das palavras que as compõem, existe um intervalo no qual se encontram aquelas expressões cujas partes composicionais ocorrem em meio a uma expressão globalmente não composicional.

Neste contexto Nunberg et al. (1994) propõem uma classificação de Ems, utilizando o conceito de decomponibilidade semântica, segundo o qual muitas expressões parecem ser composicionais semanticamente, se considerarmos que alguns dos seus componentes têm significados não-padrão. De acordo com os autores, existem três grandes grupos de EMs do ponto de vista semântico: as expressões não decomponíveis como *sinal verde* (autorização) e *dedo duro* (delator), as expressões idiossincraticamente decomponíveis como *perder as estribeiras* (perder a paciência) e *faca de dois gumes* (ação com duas consequências/interpretações opostas), e finalmente as expressões decomponíveis, como *panela de pressão* e *entre quatro paredes*.

O primeiro grupo corresponde às expressões que possuem uma interpretação atômica: não é possível, por exemplo, em *sinal verde*, definir qual das partes corresponde ao significado de *autorização*. Já as partes de uma expressão idiossincraticamente decomponível podem ser interpretadas separadamente se considerarmos que cada um dos componentes assume um sentido não convencional na expressão. Um exemplo é a expressão inglesa *spill the beans*, onde *spill* é entendida como *revelar* e *beans* como *segredos*. O mesmo ocorre, por exemplo, para a EM em português *engolir um sapo*, onde *engolir* significa *acatar* e

*sapo* pode significar *desaforo*. Dessa forma, quando se estende a semântica das partes da expressão, ela pode ser interpretada composicionalmente (*revelar os segredos*, *acatar um desaforo*). Finalmente, expressões decomponíveis são aquelas nas quais as palavras que as compõem possuem seu significado usual, porém quando usadas em conjunto, uma camada de significado adicional é criada. Isso ocorre, por exemplo, com a expressão *entre quatro paredes*, que significa *intimamente* ou *privativamente*, numa analogia às paredes de uma casa, independente de quantas sejam elas. Isto é, o sentido das palavras *entre*, *quatro* e *paredes* não é modificado pela expressão, porém quando usadas em conjunto elas assumem uma significação adicional, que complementa o sentido individual de cada uma.

Como no caso da flexibilidade sintática, Villavicencio e Copestake (2002) verificaram, utilizando essa classificação, que a grande maioria das expressões idiomáticas em uma amostra de 100 casos eram semanticamente decomponíveis (76,5% das expressões verbais).

## 2.4. Caracterização pragmática

Algumas EMs podem apresentar particularidades com relação às situações nas quais elas são empregadas. Isso significa que, por vezes, apesar de apresentarem características lexicais, sintáticas e semânticas convencionais, determinadas expressões são pragmaticamente idiossincráticas por serem inerentemente dependentes de determinada situação ou circunstância. Exemplos dessas expressões são *bom dia* e *feliz aniversário*, que são conhecidas como expressões situadas, dependentes de um determinado momento do dia ou data. Por outro lado, expressões não situadas são casos em que, apesar de existir uma dependência do contexto de uso, a situação temporal é independente, por exemplo, *first off* (*em primeiro lugar* – o uso dessa expressão depende do fato que nada foi dito anteriormente).

## 2.5. Caracterização estatística

Algumas EMs são simplesmente combinações de palavras que co-ocorrem com frequência. Essas combinações decorrem de diversos

fatores que levam uma determinada comunidade linguística a preferir determinadas combinações em detrimento de outras (históricos, políticos, etimológicos, cognitivos, etc.). Consequentemente, essas combinações assumem o estatuto de *colocações*, ou seja, maneiras convencionais de dizer determinada coisa. Muitas vezes, as colocações apresentam características convencionais, ou seja, são combinações gramaticais de palavras corriqueiras com semântica composicional, de foma que a expressão assume as propriedades das palavras que ela contém. No entanto, apesar de apresentarem semântica transparente, as colocações não aceitam substituição das palavras componentes por sinônimos ou por palavras semanticamente relacionadas, pois essas últimas tendem a parecer artificiais ou pouco naturais para um locutor nativo.

Exemplos dessas expressões são *café forte*, *chuva torrencial*, *muito obrigado*, *feijão e arroz, café com leite.* Uma forma de identificar colocações é tentar modificar uma das palavras envolvidas. Nesse caso, o resultado é o que se chama de anti-colocação, ou seja, uma combinação compreensível para um locutor nativo, porém pouco natural e por vezes inclusive cômica (Pearce, 2001). Nos exemplos anteriores, algumas anticolocações seriam *arroz e feijão, leite com café, café poderoso*, *bastante obrigado, chuva agressiva*. Siyanova e Schmitt mostram que, uma simples busca na Web (através do sistema de busca da Google) pelos termos *fish and chips* (prato típico inglês) retorna 2.090.000 páginas contra 15.900 páginas para a anti-colocação *chips and fish.* Analogamente, a expressão em português *café com leite* aparece em 632.000 páginas segundo Google, enquanto *leite com café* ocorre somente 97.700 vezes.

## *EMs em corpora especializados*

Apesar de tratar-se de construções diferentes, os processos de transformação de sentido através dos quais uma palavra ou um conjunto de palavras assumem o estatuto de *termo* e o estatuto de *Expressão Multipalavra* é análogo. Ambos são guiados por dois princípios: o da precisão e o da convenção. Na linguagem de propósito geral (em oposição àquela usada na comunicação especializada a um domínio), EMs são frequentemente empregadas para especificar uma

situação ou ação que não pode ser descrita apenas com palavras simples. Ao longo do tempo, essas formas usuais ou convencionais tornam-se cada vez mais concretas e se integram à linguagem, como se pode facilmente perceber no caso das colocações.

Analogamente, a linguagem especializada tem por particularidade um esforço da parte dos locutores no sentido de minimizar a ambiguidade, a fim de facilitar a transmissão do conhecimento. Por isso mesmo, em textos técnicos e científicos, palavras corriqueiras assumem por vezes sentidos novos, ligados a um conceito ou objeto do domínio no qual aquela comunicação se insere. Em Informática, por exemplo, o termo *árvore* designa uma estrutura de dados complexa composta por nodos e arestas ligando esses nodos de forma não cíclica. A mesma palavra, em Botânica, designa uma planta lenhosa de grande porte, e esse último sentido é o que mais se aproxima do uso dessa palavra na linguagem cotidiana. Os fatores que levam à atribuição de uma semântica nova, especializada, a uma palavra, são certamente diversos. No entanto, o mesmo princípio de convencionalidade ou até mesmo de idiossincrasia estatística pode ser usado para explicar porque um determinado termo é preferido em detrimento de outro.

Existem diversas acepções do que vem a ser um *termo*, porém sabe-se que, para especialistas de um domínio, termos são a representação do conhecimento da área específica (Krieger e Finatto, 2004). Seguindo a linha da Teoria Geral da Terminologia, acredita-se que existe uma relação 1:1 entre termos e conceitos abstratos. Essa definição, no entanto, é bastante criticada por ser reducionista e não levar em conta o termo no seu habitat natural: o texto. Independentemente da definição, termos e EMs possuem alguns aspectos semelhantes, conforme exemplificado anteriormente, no seu processo de construção do sentido. Do ponto de vista das aplicações de PLN, termos e EMs também compartilham uma característica: por apresentarem uma grande variedade em sua forma de ocorrência, ambos constituem um desafio para o tratamento uniformizado em sistemas de PLN. No que diz respeito aos termos, muitas vezes os sistemas computacionais são criados através um processo de aprendizado guiado pelos dados (*corpora*) em formato eletrônico, que

por sua vez possuem cobertura limitada em termos de representatividade. Quando um sistema construído através dessa metodologia é aplicado a um texto que pertence a um domínio diferente, muitas vezes ele apresenta dificuldades para lidar com construções desconhecidas, seja no nível léxico, sintático ou outro. A construção de recursos linguísticos, tanto para tradutores humanos quanto para sistemas de PLN, é um processo oneroso que envolve não somente o trabalho de lexicógrafos e terminógrafos como também a interação com especialistas do domínio em questão. É por isso mesmo que, recentemente, a comunidade de PLN tem dedicado grande interesse à aquisição e tratamento automáticos ou semi-automáticos de termos e EMs.

Entretanto, vale lembrar que existem diversos pontos em que o estudo de termos o estudo das EMs divergem. Em primeiro lugar, um termo não é necessariamente uma construção formada por mais de uma palavra como é o caso de uma EM. Desse fato decorre, por exemplo, que a maioria dos sistemas de PLN atuais que realizam algum tipo de tratamento especializado se concentram em termos simples e desconsideram as interações entre palavras. Outra diferença essencial, portanto, pode ser observada através dos métodos distintos desenvolvidos para identificar e processar termos e EMs. Por último, EMs estão onipresentes na linguagem cotidiana e específica, de maneira mais ou menos frequente, porém ubíqua. Apesar de existirem divergências com relação à caracterização da linguagem especializada em si, os termos e o seu estudo são próprios a ela. Até mesmo nas ciências que estudam essas estruturas, percebe-se o reflexo dessa distinção: enquanto EMs são objeto de interesse da Linguística geral, os termos são estudados por Terminólogos e Terminógrafos.

O objeto que pertence à intersecção entre termo e EM é usualmente denominado Termo Multipalavra (TM). Trata-se simplesmente de uma locução ou expressão multipalavra que possui estatuto terminológico. Destaca-se, contudo, que o mesmo não corresponde ao conceito de fraseologia do domínio. O primeiro aceita pouca variabilidade (morfológica, raramente sintática) enquanto o segundo é altamente flexível. Enquanto o primeiro representa um único conceito, não é raro que o segundo seja uma estrutura complexa que

associe mais de um conceito. As fraseologias de domínio, portanto, são um tipo de EM bastante comum nos textos especializados, porém não são estudadas no âmbito do tratamento automático de TMs.

Da mesma forma que se estima (Jackendoff, 1997) que na linguagem genérica o léxico de EMs corresponde, em número de lexemas, ao léxico das palavras simples, Krieger e Finatto (2004) assumem, com base em observações empíricas, que aproximadamente 70% das entradas em uma terminologia são TMs. Por essa razão, qualquer sistema computacional que pretende tratar textos advindos de um determinado domínio deveria possuir um dicionário contendo os TMs do mesmo. Assim, alguns trabalhos têm sido desenvolvidos nessa área para a aquisição e processamento computadorizados de TMs.

## 3. Termos multipalavras em sistemas de PLN

A extração automática ou semiautomática de termos a partir de *corpora* textuais vem atraindo o interesse de pesquisadores tanto em Linguística Computacional quanto em Terminologia há bastante tempo. Com efeito, um dos primeiros trabalhos a apresentar uma abordagem pragmática para esse problema é o trabalho de Justeson e Katz (1995). O método proposto por esses autores é relativamente simples e consiste em extrair todas as sequências que obedecem a um certo conjunto de padrões morfossintáticos para em seguida ordená-los por número de ocorrências. Além da metodologia, os autores também sugerem uma lista de sete padrões de extração para a língua inglesa, que se aplicam a todos os domínios do conhecimento:

- substantivo-substantivo (*tumor cell, ozone layer*);
- adjetivo-substantivo (*natural selection*, *inhibitory enzime*);
- adjetivo-adjetivo-substantivo (*endangered animal species, xenoreactive natural antibody*);
- adjetivo-substantivo-substantivo (*cumulative distribution function, lymphoblastoid cell line*);
- substantivo-adjetivo-substantivo (*mean squared error, colostrum inhibitory factor*);

- substantivo-substantivo-substantivo (*class probability distribution, zinc finger type*);
- substantivo-preposição-substantivo (*degree of freedom, History of Art*).

Apesar da aparente simplicidade, Manning e Schütze (1999) afirmam que esse método é bastante eficaz quando se analisa os primeiros termos extraídos, e é também bastante popular em sistemas de PLN em que uma grande revocação não é necessária. No entanto, o método não é capaz de capturar expressões contendo palavras não adjacentes, e possui um silêncio considerável com relação a todos os termos que são ignorados por serem pouco frequentes ou possuírem um padrão morfossintático diferente dos enunciados acima.

Frantzi, Ananiadou e Mima (2000) introduzem um método híbrido no qual os candidatos são extraídos através de padrões morfossintáticos rasos (como no método de Justeson e Katz) e subsequentemente submetidos a um teste estatístico chamado *C-Value*. A vantagem dessa medida é que ela é capaz de lidar com termos aninhados, ou seja, termos que contém outros termos, sendo assim capaz de extrair TMs arbitrariamente longos. Por outro lado, medidas mais sofisticadas, como aquelas baseadas em tabelas de contingência usadas na extração de EMs genéricas, poderiam eventualmente levar a melhores resultados.

Do ponto de vista da sintaxe, TMs em áreas como indústria automotiva, agricultura, química ou biomedicina tentem a seguir determinados padrões de etiquetas morfossintáticas que incluem na maioria dos casos adjetivos e substantivos. Por exemplo, em um *corpus* biomédico como o Genia, termos costumam ser nomes de doenças, substâncias, genes e proteínas (OHTA; TATEISI; KIM 2002). Assim sendo, trabalhos que descrevem métodos para a aquisição automática de substantivos compostos são também relevantes para a identificação de TMs.

Além da tarefa de identificação, existe interesse em organizar e interpretar automaticamente os termos e TMs etraídos de forma a hierarquizar os conceitos em uma ontologia ou tesauro do domínio. Nesse sentido, o trabalho de SanJuan et al. (2005) descreve a

comparação de uma ontologia construída manualmente com uma ontologia construída de forma automática sobre o *corpus* Genia. Na construção desta última, os autores mesclam três técnicas baseadas em recursos linguísticos distintos: associação estatística, WordNet e métodos de agrupamento em inteligência artificial (*clustering*).

Em particular na área da biomedicina, da medicina e da biologia, diversos trabalhos concentram-se na descrição e indexação de documentos a partir da identificação automática de termos e TMs, muitas vezes denominados de forma geral *Entidades Nomeadas*. Hagège et al. (2002) descrevem um sistema baseado em regras lexicais e sintáticas para a identificação de genes e proteínas. Dada a disponibilidade de recursos como dicionários e ontologias nesse domínio, em especial em língua inglesa, não são raros os sistemas que realizam uma simples busca de padrões (gerados a partir das entradas do dicionário) no *corpus* em questão. Para esclarecer as particularidades relacionadas a esse domínio, mas mais amplamente ao trabalho de identificação e tratamento de TMs em *corpora* de domínio, apresenta-se na próxima seção um estudo de caso realizado sobre o *corpus* Genia no sentido de caracterizar os TMs nele encontrados.

## 4. Estudo de caso: TMs em biomedicina

A fim de exemplificar as EMs de domínio, iremos analisar os TMs em um *corpus* da área biomédica. Trata-se do *corpus Genia*, um conjunto de dois mil resumos de artigos científicos da área biomédica extraídos da base de artigos Medline com base nas palavras-chave “humano”, “célula sanguínea” e “fator de transcrição” (OHTA; TATEISI; KIM 2002). Ele contém 18,5 mil frases e 490,7 mil palavras, anotados com informação de classe morfossintática. Além disso, os resumos também foram anotados por especialistas com relação a uma terminologia do domínio. Que contém conceitos como nomes de doenças (tumor fibroblástico) e células (linfócito T primário).

Tabela 3 – Dimensões do *corpus* de biomedicina analisado (Genia).

| | |
|---|---|
| Resumos | 2000 |
| Frases | 18519 |
| Palavras (tokens) | 490752 |
| Número médio de palavras por frase | 26,5 |
| Termos anotados | 97876 |
| TMs anotados | 55487 |
| Número médio de termos por frase | 5,29 |
| Número médio de TMs por frase | 3 |
| Número médio de palavras por termo | 2,06 tokens |
| Número médio de palavras por TM | 2,86 tokens |

A Tabela 3 resume algumas das dimensões do *corpus*, que permitem uma análise superficial dos seus TMs. Em primeiro lugar, nota-se que, dos 97876 termos anotados no *corpus*, 55487 são multipalavras, ou seja, em torno de 57%, em conformidade com as estimativas apresentadas anteriormente (entre 50% e 70%). É igualmente interessante perceber que o número médio de palavras por frase é superior ao que se espera em um *corpus* genérico de língua inglesa (que fica em torno de 22 palavras). Além disso, essas frases contém um número surpreendentemente grande de termos e de TMs: cada frase tem em média 5,29 termos, dos quais 3 possuem mais de uma palavra. Na tentativa de caracterizar esses termos de acordo com o número de palavras, percebe-se que se os termos são considerados de forma homogênea, têm um comprimento médio de 2 palavras, enquanto que os TMs têm um comprimento médio de quase 3 palavras. Isso significa que não somente os TMs são muito presentes nesse *corpus*, como também eles são relativamente longos (podendo chegar a termos anotados com 22 palavras). Uma explicação possível para essa onipresença dos TMs é o aninhamento de termos.

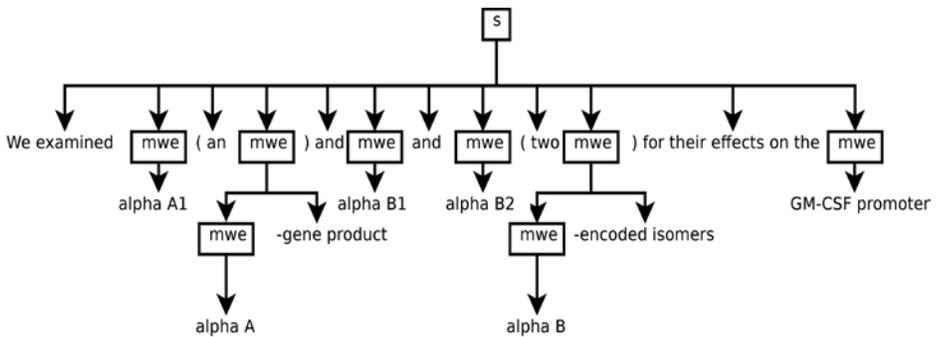



Figura 1 – Exemplo de frase extraída do *corpus* Genia e sua anotação de termos.

O exemplo apresentado na Figura 1 ilustra bem esse aninhamento, por exemplo, o termo *alpha A* está contido na expressão *alpha A-gene product*. Da mesma forma, *alpha B-encoded isomers* é um TM aninhado que contém outro TM. Dessa forma, a frase acima contém oito TMs, o que explica de alguma forma o grande número de TMs por frase. Por outro lado, o uso da terminologia em qualquer domínio é outro fator que se reflete nessa dimensão de análise.

Dessa forma, uma rápida análise para caracterizar o *corpus* mostra alguns dos desafios que devem resolver não somente terminógrafos e lexicógrafos mas também sistemas de PLN que realizam a identificação de TMs e EMs em domínios especializados. Além da questão da onipresença e grande número de construções multipalavras, adicionam-se tantos outros como a normalização da grafia (em biomedicina, alguns autores hifenizam *T-cell* enquanto outros separam as palavras *T cell*), o uso massivo de acrônimos, abreviaturas e elipses (*HIV1, IL2*). Finalmente, existem ainda aqueles problemas relacionados com o próprio tratamento linguístico-computacional dos textos, sejam eles de domínio ou genéricos, como a resolução da ambiguidade semântica e/ou sintática, a interpretação de anáforas e correferências, a análise sintática, etc. Em resumo, um dos grandes desafios na área de PLN consiste na adaptação de sistemas e recursos para tratar um domínio ou domínios especializados e para auxiliar especialistas a automatizar parte do trabalho de criação de recursos linguísticos como as terminologias.

## Referências

Dimitra Anastasiou, Chikara Hashimoto, Preslav Nakov, and Su Nam Kim (eds). *ACL/IJCNLP 2009 Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications*, Singapore.

Timothy Baldwin. The deep lexical acquisition of English verb-particles. *Computer Speech and Language, Special Issue on Multiword Expressions*, 19(4): 398–414, 2005.

Timothy Baldwin. Compositionality and Multiword Expressions: Six of One, Half a Dozen of the Other? Invited Talk. In Proceedings of the *ACL-2006 Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, Sydney, Australia, 2006.

Cristina Cacciari. A psycholinguistic point of view: are idioms really relevant for theories of language comprehension and production? Presented at *Collocations and idioms 2003: linguistic, computational, and psycholinguistic perspectives*. 2003.

Nicoletta Calzolari, Charles Fillmore, Ralph Grishman, Nancy Ide, Alessandro Lenci, Catherine MacLeod, and Antonio Zampolli. Towards best practice for multiword expressions in computational lexicons. In Proceedings of the *3rd International Conference on Language Resources and Evaluation* (LREC 2002), pages 1934–1940, Las Palmas, Canary Islands, 2002.

Gael Dias, Sylvie Guilloré and José Pereira Lopes. Multilingual Aspects of Multiword Lexical Units. In *Proceedings of Workshop on Language Technologies,* Ljubljana, Slovenia, pages 11—21, 1999.

Charles Fillmore, Paul Kay, and Mary C. O'Connor. Regularity and idiomaticity in grammatical constructions. *Language*, 64:501–538, 1988.

Charles Fillmore. Multiword Expressions: An Extremist Approach. Presented at *Collocations and idioms 2003: linguistic, computational, and psycholinguistic perspectives*. 2003.

Alexandre Gil and Gael Dias. Efficient Mining of Textual Associations. In Proceedings of the *International Conference on Natural Language Processing and Knowledge Engineering*, Beijing. China, pages 549—555, 2003.

Caroline Hagège, Agnes Sándor, and Anne Schiller. *Linguistic processing of biomedical texts.* In Elisabete Ranchod and Nuno J.

Mamede, editors, PorTAL, volume 2389 of LNCS, pages 197–208, Faro, Portugal, June, Springer, 2002.

*Ray Jackendoff. The Architecture of the Language Faculty. MIT Press, Cambridge, USA, 1997.*

John S. Justeson and Slava M. Katz. Technical terminology: Some linguistic properties and an algorithm for identification in text. *Natural Language Engineering*, 1(1):9–27, 1995.

Graham Katz and Eugenie Giesbrecht. Automatic identification of non-compositional multi-word expressions using latent semantic analysis. In Proceedings of the *ACL-2006 Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pages 28–35, Sydney, Australia, 2006.

Anna Korhonen. Current trends and future challenges in computational linguistic research on multiword expressions. *Presented at Collocations and idioms 2006: linguistic, computational, and psycholinguistic perspectives*. 2006.

Chris Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing*, MIT Press. Cambridge, MA, 1999.

Diana McCarthy, Bill Keller, and John Carroll. Detecting a continuum of compositionality in phrasal verbs. In Proceedings of the *ACL2003 Workshop on Multiword Expressions: analysis, acquisition and treatment*, pages 73–80, Sapporo, Japan, 2003.

Diana McCarthy. Automatic methods for detecting the compositionality of multiwords. Presented at *Collocations and idioms 2006: linguistic, computational, and psycholinguistic perspectives*. 2006.

Rosamund E. Moon. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford University Press, Oxford, UK, 1998.

Geoffrey Nunberg, Ivan A. Sag, and Tom Wasow. Idioms. *Language*, 70:491–538, 1994.

Andrew Pawley. Developments in the study of formulaic language since 1970: A personal view. In Paul Skandera (ed.) *Phraseology and Culture in English*, pages 3–48 Mouton de Gruyter, 2007.

Darren Pearce. Synonymy in collocation extraction. In Proceedings of the *NAACL 2001 Workshop on WordNet and Other Lexical Resources:Applications, Extensions and Customizations*, pages 41–46, Pittsburgh, USA, 2001.

Susanne Riehemann. *A Constructional Approach to Idioms and Word Formation*. PhD thesis, Stanford University, 2001.

Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. Multiword expressions: A pain in the neck for NLP. In Proceedings of the *3rd International Conference on Intelligent Text Processing and Computational Linguistics* (CICLing-2002), pages 1–15, Mexico City, Mexico, 2002.

Antoine Tremblay, Bruce Derwing, Gary Libben and Chris Westbury. Are Lexical Bundles Stored and Processed as Single Units? In Proceedings of the *25th UWM Linguistics Symposium on Formulaic Language*. Milwaukee, Wisconsin, 2007.

Aline Villavicencio and Ann Copestake. On the nature of idioms. *LinGO Working Paper* No. 2002-01. 2002.

Alison Wray. Formulaic language in learners and native speakers. *Language Teaching*, 32:213—231, 1999.

Alison Wray. *Formulaic Language and the Lexicon*. Cambridge:CUP, 2002.

# O TEXTO JURÍDICO EM TRADUÇÃO E VERSÃO: PROBLEMAS E SOLUÇÕES NA LINGUAGEM DOS CONTRATOS

*Cristina Lopes Perna*[1]

**Resumo**: *A linguagem jurídica apresenta peculiaridades entre textos jurídicos de duas culturas, dado o seu caráter técnico e específico. A tradução de um texto jurídico de um idioma a outro resulta em tarefa árdua para o tradutor, visto que os documentos legais exigem um tipo especial de tradução e o tradutor desse tipo de linguagem especializada fica mais restrito do que em outros tipos de textos. Ao traduzir um texto jurídico, o tradutor deve manter a precisão deste, encontrando equivalentes legais que em sua relevância, correspondam tanto ao texto original da língua fonte como ao texto traduzido da língua alvo. É indiscutível que há conceitos culturais que designam elementos peculiares a cada língua. Essa crença nos permite afirmar que é inevitável certa perda de sentido no processo de tradução de textos jurídicos, especialmente por pertencerem a diferentes sistemas legais. No caso do presente texto, discutiremos as dificuldades em se traduzir e verter contratos internacionais em português- inglês-português, visto que os sistemas jurídicos do Brasil e EUA são de naturezas diferentes, a saber, o brasileiro é originário do Civil Law ao passo que o Estadunidense, do Common Law.*

## 1. Textos jurídicos: sua natureza e peculiaridades

Não há dúvida de que há conceitos culturais específicos para designar elementos peculiares de cada língua. Esta noção data dos estudos de Sapir, estudioso que afirmou que “não há semelhança suficiente entre duas línguas que nos permita considerá-las como representantes da mesma realidade social (1956:69)”. Portanto, é de se esperar que uma perda considerável de significado seja inevitável

[1] Doutora em Letras. Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) – Faculdade de Letras - Endereço eletrônico: cperna@pucrs.br.

durante o processo tradutório de textos jurídicos, especialmente porque diferentes culturas apresentam sistemas jurídicos diferentes. Vários estudiosos de tradução jurídica demonstraram a dificuldade em se garantir o equivalente terminológico mais próximo do texto fonte (doravante TF) no texto alvo (doravante TA).

Newmark (1981) argumentou que textos jurídicos exigem um tipo especial de tradução, pela simples razão de que o tradutor deste tipo de linguagem especializada fica mais restrito do que ao traduzir outro tipo de gênero. Schwartz (1977) chamou atenção para o fato de que quando traduzimos textos jurídicos, a principal tarefa do tradutor é garantir que a tradução seja feita da forma mais precisa possível, procurando encontrar equivalentes linguísticos que, em sua relevância jurídica, tenham correspondência tanto no TF como no TA.

Não obstante, Sarčevic (1997) afirma que os tradutores jurídicos não podem ter expectativas de usar equivalentes naturais do sistema jurídico do TA que sejam idênticos aos do TF, pois ocorrem "incongruências inerentes de terminologia devido aos diferentes sistemas jurídicos" (1997:235). Entretanto, a autora argumenta que é perfeitamente legítimo que se exija que estes tradutores utilizem "o equivalente mais próximo" do sistema jurídico do TA, em outras palavras, o termo que mais proximamente transmita o sentido jurídico do TF. A autora também acrescenta que esta é uma tarefa muito árdua.

Obenaus (1995) alega que, embora a equivalência lexical pudesse ser mantida e que é possível designarem-se vários graus de equivalência. Os linguistas sabem que uma correspondência de um a um entre termos de línguas diferentes é a exceção e não a regra. Malmkjær (1998) admite que seja possível se obter uma equivalência entre TF e TA em um ou mais níveis distintos, porém, segundo a autora, é impossível que se faça isso, considerando o texto em sua totalidade.

A dificuldade em se manter a equivalência em tradução jurídica reside no fato de que os sistemas jurídicos são diferentes. Estes sistemas são geralmente classificados em grupos ou famílias. Há dois grupos principais, dependendo da tradição: o sistema *common law* e o sistema *civil law*. *Common law* (em português, "lei comum") é um sistema legal oriundo da Inglaterra, utilizado lá e na maioria dos países

que foram colônias ou territórios britânicos (Perna, 2010). Sua característica principal é a valorização da jurisprudência em detrimento das leis estatutárias. *Civil law* não é o que denominamos de Direito Civil. No sistema jurídico que adota o *civil law*, a legislação representa a principal fonte do Direito. Os tribunais fundamentam as sentenças nas disposições de códigos e leis, a partir dos quais se originam as soluções para cada caso. O sistema legal da Grã-Bretanha, dos EUA e da Nova Zelândia, entre outros, pertencem ao primeiro. O segundo, que se originou da lei romana, é o sistema jurídico da Alemanha, Áustria e Brasil, entre outros. É a lei imposta pelo estado, baseada em normas estabelecidas pelo poder legislativo.

Exemplos de incongruências de termos jurídicos advindos de diferentes culturas foram observados primeiramente na área de direito comparativo. A necessidade de se comparar leis de diferentes países originou do fato de que, para criar leis nacionais, era necessário observar códigos estrangeiros. A área de Estudos da Tradução beneficiou-se dos estudos do direito comparativo assim como da área de comércio exterior, pois estas vêm fornecendo subsídios para a identificação das diferenças jurídicas e, portanto, de termos jurídicos que não são equivalentes. Obenaus (1995) chamou atenção para o fato de que mesmo um termo que havia sido considerado como tendo um sentido universal como "juiz", deve ser visto a partir de outra perspectiva quando traduzido, visto que em um país que segue o sistema legal da *common law,* um juiz tem menos poder do que um juiz de um país que obedece ao sistema de *civil law.* Em outras palavras, no primeiro o juiz não pode chegar a um veredicto, o qual é obtido através do júri, enquanto que no segundo, é o juiz que chega a um veredicto, na maioria dos casos.

No passado, houve alguns esforços no sentido de aproximar as lacunas entre os dois sistemas jurídicos, graças a uma aceitação de tradução menos literal. Este fato ainda vem ocorrendo no Canadá, um dos poucos países que é não somente bilíngue, como também bilegal, conforme observou Sarčevic (1998). Obviamente, a dificuldade lá reside no conflito entre o sistema *common law*, das províncias anglófonas, e o sistema *civil law*, da província de Quebec.

Também no Brasil, como em muitos outros países, a tradução de termos jurídicos apresenta sérias dificuldades para o tradutor. Há um número de conceitos que possuem somente equivalência parcial, se compararmos os termos jurídicos entre a língua inglesa e a portuguesa e que podem causar ambiguidade no TA. Theodoro Jr. (1999) citou uma série de exemplos nos quais um termo do TF não pode ser necessariamente substituído por outro no TA. Para ilustrar, vejamos o termo *joint venture*, definido no *Black's Law Dictionary* como "*a business undertaking by two or more persons engaged in a single defined project*"[2]. Segundo Theodoro Jr., encontram-se algumas traduções de *joint venture* como "empreendimento conjunto" o que, segundo o autor, significa "conduzir um negócio junto". Danilo Nogueira (2000) chama atenção dos tradutores quanto à tradução de *joint venture* por "sociedade em conta de participação", cujo equivalente em inglês é *joint adventure.*

Theodoro Jr. segue sua argumentação apresentando áreas problemáticas, nas quais o tradutor poderá encontrar desafios em relação aos sistemas jurídicos brasileiros e estadunidenses. De acordo com o autor, o termo *oficial de justiça* foi vertido para o inglês de várias maneiras, que vão desde *sheriff* ou *bailiff* até *judicial officer*, visto que o tradutor não encontrou uma correspondência no sistema jurídico do TA. No sistema jurídico brasileiro, este servidor é responsável por trazer um certificado de serviço, enquanto que no sistema estadunidense a atribuição do *bailiff* é manter a ordem durante os processos judiciais, de acordo com o *Black's Law Dictionary.*

Estes são apenas alguns exemplos de incongruências entre termos jurídicos que podem acarretar em grandes desafios para os tradutores. Esforços têm sido feitos por parte de tradutores jurídicos na tentativa de reduzir estes desafios (Weston, 1983 e Sarčevic, 1985), porém soluções eficazes para esta situação ainda não foram encontradas, com exceção de algumas poucas estratégias apresentadas por tradutores clássicos (ver Vinay e Darbelnet, 1958; Wotjak, 1981; Malone, 1988 e Newmark, 1988), para tentar facilitar a ausência de equivalência em textos gerais. Para Deslile (1988:72), "os procedimentos tradutórios de Vinay &

[2] Um empreendimento no qual uma ou mais pessoas estão envolvidas em um único e definido projeto [minha tradução].

Dalbernet não ajudam o tradutor a encontrar equivalentes para a tradução. Um procedimento é um método para se obter resultado, (...) para se conduzir uma atividade até sua conclusão. Porém esses 'procedimentos' são, na verdade, fatos atrelados a resultados".[3]

Sarčevic (1985) apresenta seus procedimentos a partir de uma perspectiva diferente. A autora preocupa-se com o fato de que a maioria das traduções jurídicas sejam originalmente escritas privilegiando os interesses do país da língua fonte, de modo que o tradutor, segundo ela, deve aderir ao TF. Devido a esta aderência, a autora afirma que os tradutores tendem a favorecer o que ela chama de "procedimentos tradutórios orientados à língua fonte", ou seja, estratégias denominadas de **transcrição** e **tradução de empréstimo**.

Para Sarčevic, **transcrição** é o procedimento através do qual o tradutor repete o termo do TF como se fosse um empréstimo ou transfere-o para a língua alvo, colocando-o em itálico. Segundo a autora, isto só é aceitável se o termo na língua fonte já tiver sido naturalizado na terminologia da língua alvo. Em relação à **tradução de empréstimo**, este é um recurso no qual o tradutor utiliza um termo composto ou expressão do TF e o traduz literalmente para o TA. Toda vez que o termo na língua fonte for semanticamente transparente, haverá uma tendência a se usar palavra-por-palavra; entretanto, ela alerta para que o tradutor se assegure de que esta estratégia não resulte em um termo já existente na cultura alvo, o que neste caso poderá se tornar um falso cognato. Segundo a autora, "... já que precisão é de importância primária na tradução jurídica, por vezes uma "tradução de empréstimo" pode ser preferível a um equivalente linguístico na língua alvo, que falha em transmitir o sentido da informação do termo na língua fonte" (1985:129).[4]

---

[3] "Vinay and Darbelnet's translation procedures do not help the translator to find translation equivalents. A procedure is a method to obtain a result, (…) of carrying an activity through to its conclusion. But these 'procedures' are in fact labels attached to results (Deslile, 1988:72)".

[4] "Since precision is of primary importance in the translation of laws, sometimes a loan translation may be preferable to a potential target language equivalent that fails to convey the significance of the information content of the source language term" (1985:129).

Não obstante, Sarčevic argumenta que tradução de empréstimo, no caso de termos específicos a uma cultura, pode resultar em incompreensão. Sua principal preocupação é determinar até que ponto o tradutor jurídico deve fazer concessões ao leitor, ao usar o que ela chama de “procedimentos de tradução orientados à língua alvo”. Ela os divide em três categorias: (1) adaptação, (2) descrição através de definições e explicações e (3) substituição descritiva.

**Adaptação** significa utilizar um equivalente cultural no TA ao invés daquele usado no TF. Visto que a tradução jurídica deve garantir equivalência de significado, Sarčevic argumenta que a adaptação é “duvidosa e deve ser evitada na tradução jurídica" (1985:131). Entretanto, Sarčevic não sugere o que fazer em casos nos quais a forma rígida do documento demanda que a adaptação de certas expressões seja feita no TA, como no caso de documentos como contratos, cujas partes são formadas basicamente por expressões formulaicas. Por exemplo, o tradutor deverá atentar para termos como *whereas*, se estiver vertendo um contrato para o inglês. Segundo a convenção de contratos no Brasil, *whereas* não deve ser traduzido pelo seu equivalente no português *enquanto*, e sim por *considerando que*.

O segundo tipo de procedimento considerado como orientado ao TA é aquele que Sarčevic chama de **definições e explanações**. De acordo com a autora, este é um tipo de *overtranslation*, frequentemente empregado por tradutores jurídicos quando há muitos termos específicos da cultura fonte. Este procedimento é feito por intermédio de glossários, prefácios, notas de rodapé ou comentários.

O terceiro tipo é chamado de **substituto descritivo**, que descreve a forma e/ou função da instituição ou conceito em questão; entretanto, este não é um equivalente cultural. Neste caso, o tradutor deve utilizar um *translating couplet*, em outras palavras, uma **tradução de empréstimo** reconhecida, seguida pelo termo do TF em parênteses. De acordo com Sarčevic, "no restante do texto a **tradução de empréstimo** é geralmente citada, por exemplo *work organization* (*firm*), citado *a posteriore* como *work organization*" (1985:131).

Sarčevic afirma que "não basta que os tradutores meramente identifiquem os equivalentes mais próximos nos sistemas jurídicos fonte e alvo" (1997:149). O que os tradutores devem buscar como

equivalência é uma perspectiva pragmática de equivalência, na qual o conteúdo proposicional[5] do TF deve estar presente no TA. Para atingir este fim, o tradutor deverá procurar uma compensação para quaisquer diferenças conceituais que podem vir a colocar em risco a interpretação e aplicação do conteúdo proposicional.

## 2. A tradução de contratos internacionais

Ao contrário de traduções literárias ou mesmo traduções técnicas, que possuem autoria e leitores, os contratos internacionais não seguem esta regra. O objetivo principal de um contrato é selar um acordo entre duas partes, neste caso cada parte pertence a uma língua e contexto; portanto, a tradução deve ser construída em duas vias – ambas as partes estão comprometidas.

De acordo com o *Duhaime's Law Dictionary*[6], a lei de contratos é originária da *common law*, que é posterior à *tort law*[7]. Tanto os contratos como os *torts* levam a obrigações. Os contratos são veículos através dos quais as pessoas voluntariamente criam obrigações para si mesmas. As promessas são a base dos contratos. Um contrato é feito de promessas de uma pessoa (ou partes, ou empresa) para que se cumpra algo em troca de uma promessa de outra pessoa. A lei de contrato existe para assegurar que se cumpram as promessas feitas e, caso contrário, a lei os obrigará a tal.

Quanto à forma e estilo, os contratos são de certa forma padronizados, o que os torna mais fáceis de serem traduzidos, pois o tradutor se familiariza com o formato. As orações presentes nos contratos são chamadas pela literatura especializada de *boiler plates*

---

[5] Dizer que uma crença é uma representação é simplesmente dizer que ela tem um conteúdo proposicional e um modo psicológico, que o seu conteúdo determina um conjunto de condições de satisfação sob certos aspectos, que o seu modo psicológico determina uma direção de ajustamento do seu conteúdo proposicional, de tal maneira que todas estas noções – conteúdo proposicional, direção de ajustamento, etc. – são explicadas pela teoria dos atos de fala. (Searle, 1983: 12)

[6] www.duhaime.org/LegalDictionary.aspx

[7] cf. **Black´s Law Dictionary,** *Tort Law* é um ramo da lei que lida com problemas civis para os quais uma solução deve ser obtida.

ou expressões formulaicas, que incluem expressões como “força maior” e “leis aplicáveis”. Entretanto, além destas expressões padrões, um contrato pode conter uma linguagem bastante específica, dependendo do tipo de acordo que está sendo feito. Se for um contrato comercial de distribuição de certo produto, o tradutor deve também se familiarizar com a linguagem especializada para aquele produto além, é claro, de possuir um excelente comando da língua do TA e TF.

A fim de garantir requerimentos evidenciais, quando se compara a lei contratual entre contratos brasileiros e estadunidenses, algumas discrepâncias aparecem de início. Os contratos brasileiros possuem preâmbulos que dão detalhes sobre as partes e as razões para selarem o acordo. Por exemplo, ao descreverem as partes[8], o contrato brasileiro incluirá informação sobre a nacionalidade da pessoa, estado civil, ocupação, endereço, número de identidade e CPF. No caso das partes serem empresas, ele incluirá informação sobre o tipo de empresa, onde é sua sede, qual seu CNPQ e quem irá representá-la.

Em contrapartida, muito pouco é informado sobre as partes no contrato estadunidense, pelo simples fato de que os detalhes serão especificados quando necessário ou em respeito ao direito de confidencialidade do cliente. Não obstante, isto parecerá muito geral aos olhos do advogado brasileiro, que espera esta informação a fim de garantir a fidedignidade das partes.

É também possível de se observar diferenças entre os dois sistemas jurídicos quanto ao tipo de organização de uma empresa. Uma rápida verificada no site da Embaixada Brasileira em Washington D.C.[9] revelará que as traduções dos tipos de empresa não oferecem equivalência nas duas culturas e sim mencionam que *A closely resembles B*[10], como em “Sociedade Anônima”, cuja definição o site descreve como "*closely resembling*" uma *joint-stock company or corporation* e “Sociedade Limitada”, como “sendo semelhante” a *limited liability quota company.*

---

[8] In case the parties are individuals and not corporations.
[9] Veja o site www.brasilemb.org
[10] A se assemelha a B.

Outro problema ao traduzir preâmbulos está em definir os termos para as posições ou atribuições de certa posição na empresa. Por exemplo, em algumas situações uma pessoa está representando a Diretoria no ato, cujo termo em inglês pode ser *Board of Directors*. O problema reside no fato de que no Brasil um "diretor" é alguém que está presente na empresa todos os dias, portanto a tradução direta para *director* não é apropriada. Nos EUA, são os *officers* que estão mais familiarizados com as atividades diárias da empresa e não os *directors*, portanto a tradução para "diretor" deveria ser *officer* e para *Board of Directors*, ao invés de se traduzir para "Diretoria", ou mesmo "Quadro de Diretores", a melhor opção seria "Conselho Administrativo".

Embora haja discrepâncias culturais entre conceitos legais e categorias nos diferentes sistemas jurídicos, tem havido esforços para unificar os documentos no seu formato no direito contratual internacional[11]. Estes esforços têm sido eficazes em parte; no entanto, ocorrem ainda situações que podem ser desconcertantes para um tradutor, como no caso de algumas expressões formulaicas em contratos, visto que estas podem ser ambíguas em certos contextos e, portanto, dificultarem a transmissão correta do conteúdo proposicional do TF para o TA.

Uma das soluções para se diminuir as discrepâncias entre os conteúdos proposicionais do TF para o TA pode ser através do uso de estratégias de estreitamento e alargamento de conceito, conforme sugerido por Carston (1996). Sequeiros (2001), ao aplicar as noções de Carston, afirma que as discrepâncias originadas na tradução podem ser resolvidas se o tradutor escolher codificar sua interpretação do conceito fonte (a proposição do original).

**Estreitamento de conceito**, que também pode ser denominado de **enriquecimento**, pode ser entendido como a aplicação de um conceito que descreve uma extensão de entidades em geral para designar um conjunto mais restrito de entidades. Quando enunciamos certo conjunto de domínios, na verdade estamos nos referindo a um

---

[11] Veja *Unidroit Principles* – International Institute for the Unification of Private Law. http://www.unidroit.org/english/principles/contects.htm

subconjunto deste domínio. Podemos ilustrar este fenômeno através do exemplo (1), de Carston (1996):

(1) Ela veste coelho.

O tradutor naturalmente reduzirá o sentido do substantivo coelho, que inclui o sentido 'coisas feitas de coelho' e o enriquecerá para incluir 'pele de coelho'. Portanto, de acordo com Sequeiros, "*the use of a concept picks out only those entities to which it is relevant to apply it*" (2001:198)[12]. Este processo de estreitamento é visto no processo tradutório quando o tradutor precisa enriquecer o conteúdo linguístico do TF para transmitir o conteúdo proposicional completo da mensagem no TA, caso contrário a tradução não será bem sucedida.

O processo de **alargamento de conceito** é oposto ao anterior e pode ser definido como o uso de um conceito que é menos literal, i.e. que compartilha algumas implicações lógicas e contextuais com o original. Este é o caso de usos metafóricos como em (2):

(2) Aquela criança é realmente uma peste.

No caso de "peste", o traço semântico de "epidemia, pestilência" seria descartado, ao passo que o traço "pessoa má, perniciosa", que pertence a seus traços semânticos, permaneceria. Segundo Sequeiros, usos frouxos de conceitos podem dar margem a discrepâncias entre o TF e TA.

Para um exemplo de termo jurídico que é estreito em português, porém mais largo em inglês considere-se o caso do termo *prevarication.* Segundo o Black´s Law Dictionary, *prevarication* é "o ato ou instância de mentir ou evitar a verdade; subterfúgio". Por outro lado, em Português o termo "prevaricação" implica em um conceito bastante estreito, pois descreve um crime específico, i.e., "ato perpetrado por funcionário publico e que consiste em retardar ou deixar de praticar ato de ofício, ou em praticá-lo contra disposição legal expressa, para

---

[12] O *uso* de um conceito seleciona somente aquelas entidades para as quais é relevante aplicá-la.

satisfação de interesse ou sentimento pessoal", segundo o Novo Dicionário Aurélio. Ao verter do inglês para o português o tradutor terá de estreitar o conceito a fim de incluir somente o que se aplica ao conceito em português.

## 3. Análise e discussão

A presente análise será baseada nos *corpora* de textos jurídicos, incluindo Contratos de Licença, Acordos, Contrato para a Transferência de Unidade Industrial, Acordo de Parceria, Contrato de Vendas e Contrato de Vendas de Imóveis. Estes *corpora* são compostos de contratos tanto no português do Brasil como no inglês dos EUA, com suas respectivas traduções.

A primeira análise será no nível da oração, quando demonstrarei que as expressões formulaicas (*boiler-plate formats*) não são tão transparentes de uma cultura a outra. Como primeiro exemplo, consideremos uma oração em inglês, tirada de um Contrato de Compra de Imóveis (3):

(3) This Agreement shall be binding upon and inure to the benefit of the parties hereto and their respective successors and permitted assigns.

Todos os conceitos em (3) estão também presentes na cultura brasileira, portanto a versão para o português não apresenta grandes dificuldades. O tradutor poderá optar pela orientação ao TA e traduzir de forma direta, como em (4):

(4) Este contrato constituirá obrigações e terá efeito legal para o benefício das partes aqui citadas e de seus respectivos sucessores e cessionários legais.

Entretanto, o formato *boiler plate* na cultura alvo para a expressão *to be binding upon* é simplesmente "obrigará" e para a expressão *to inure to the benefit of* é "beneficiará". Ao analisarmos a maioria das expressões nos contratos do português do Brasil, observaremos o seguinte (5):

(5) Este contrato obrigará e beneficiará (...)

Alguns tradutores poderão arguir que, ao escolher o formato (3), o segmento será muito reduzido no TA (cf. Carson). A solução poderia ser obtida ao analisarmos as **intensões** semânticas[13] da expressão em inglês. De acordo com o **Black´s Law Dictionary**, "***to bind*** *is to impose one or more legal duties on a person or institution as in < the contract binds the parties>*[14]*"* "Impor obrigações legais" tem o mesmo efeito legal de "**obrigar**". Quanto à expressão *to inure to***,** o *Merriam Webster Unabridged Dictionary*[15] oferece uma definição que se aproxima do contexto em (3): *to become legally effective* como em < *there is such an identity of interest between the taxpayers that a refund to one will inure to the benefit of the other*>. A expressão *to become legally effective* expressa que haverá algumas condições jurídicas que garantirão o benefício das partes, portanto, neste contexto, se o tradutor optar por verter *The contract shall...inure to the benefit of the parties* para "O contrato...beneficiará", este não estará transmitindo o mesmo conteúdo proposicional do TF.

No primeiro caso, no qual o termo *shall be binding* pode ser vertido para "obrigará", não há discrepância no TF e TA, visto que ambas as expressões possuem o mesmo conteúdo proposicional. No entanto, no segundo caso, a expressão *to inure to the benefit of the parties* possui um escopo semântico maior do que sua contrapartida em português "**beneficiará"**, portanto a perda em conteúdo proposicional no TA poderá causar uma perda na força jurídica.

Constatamos, então, que se analisarmos um *corpus* através da perspectiva terminológica, é possível observar que o tradutor encontrará desafios quanto à escolha de termos, devido ao seu grau de

---

[13] According to Lyons concepts have intensions and extensions. Intensions are "the set of essential properties which determines the applicability of the term (1977:159) and extension is "the class of the things to which it is correctly applied (1977:158).

[14] Obrigar é impor um ou mais deveres legais a uma pessoa ou instituição, como em <o contrato obriga as partes>.

[15] MERRIAM-WEBSTER UNABRIDGED (www.Merriam-WebsterUnabridged.com) copyright © 2005 by Merriam-Webster, Incorporated.

ambiguidade em certos contextos. Como tais expressões são muito numerosas para criarmos aqui uma lista exaustiva, alguns exemplos devem ser suficientes para ilustrar minha posição. Nestes exemplos, utilizarei as noções de **estreitamento** e **alargamento de conceitos**, a fim de demonstrar as discrepâncias encontradas nas linguagens jurídicas de contratos entre o português do Brasil e o inglês dos EUA, tanto no TA como no TF.

Como primeiro exemplo de item lexical, consideremos o termo *covenant*. Na linguagem de contratos, a palavra *covenant* pode ser traduzida por “acordo”, “contrato”, “promessa” ou “garantia”, cujos conceitos compartilham propriedades essenciais (intensões) ou como “cláusula”. Em inglês, o conceito também possui outros dois sentidos: (1) “um acordo formal ou promessa, geralmente presente em um contrato" ou (2) "cláusula" (cf. Blacks Law Dictionary). Podemos observar estas duas situações na linguagem contratual ilustrada abaixo:

(6) *Share purchase and sale agreement and other* ***covenants****.*[16]

(7) *NOW, THEREFORE, in consideration of the mutual* ***covenants*** *and agreements set below in this Agreement, the parties hereto agree as follows*[17]*:*

No exemplo (6) parece inquestionável que a expressão *covenant* está se referindo ao instrumento em sua totalidade, visto que ela se refere ao tipo de contrato e não a partes específicas deste. A versão para o português neste caso será como em (8):

(8) Contrato de compra e venda de ações e outros **acordos**.

No entanto em (7), não está claro se a palavra *covenant* está se referindo ao contrato no todo ou somente a suas partes, pois o tradutor optou pela seguinte versão:

[16] Contrato de compra e venda de ações e outros acordos.
[17] Para tradução, ver 10.

(9) ISTO POSTO, em contraprestação às **promessas** e acordos mútuos no presente contrato, as partes do presente têm entre si justo e contratado o seguinte:

Ao optar pelo termo “promessas” (9) para verter o termo *covenants* (7), o tradutor falhou em observar que a expressão *set below*, delimita o escopo do termo para o que se segue no texto que, neste caso, é uma listagem de cláusulas; neste caso, o tradutor deveria ter optado pela palavra “cláusulas*” (clauses)* ao invés de “promessas*” (promises)*.

Uma segunda questão a ser considerada aqui é a análise de todo o enunciado (7) e sua versão (9). O tradutor optou por usar a expressão formulaica do TA em (9) ao invés de utilizar a forma mais direta, como em (10):

(10) AGORA, PORTANTO, em consideração às mútuas cláusulas e acordos abaixo estabelecidos no presente Contrato, as partes aqui contratam o que segue:

Na tentativa de adaptar ao TA e utilizar um formato convencional dos contratos em português, o tradutor não somente omitiu a tradução da expressão *set below*, o que provavelmente levou-o a traduzir “promessa” por *covenant*, como também tornou a versão mais longa no português, ao escolher a expressão “as partes do presente têm entre si justo e contratado o seguinte”, o que em retro-traducao[18] seria (11):

(11) *[t]he parties hereto believe as fair and have mutually agreed the following:*

Ao invés de “as partes aqui contratam o que segue”, como em (12):

(12) [t]he parties hereto agree as follows:

---

[18] Retro-traducao é uma técnica sugerida por Mona Baker (1992) para testar a qualidade da tradução.

Em outras palavras, como a tradução literal é também aceitável na lingaugem brasileira de contratos, não havia necessidade de verter para o português através de um enunciado mais longo a fim de assemelhá-lo ao formato *boiler plate* do TA.

Além de problemas de polissemia lexical como o descrito acima, há outros de polissemia no nível do sintagma preposicional. Por exemplo, *as of* + data é uma expressão comum em inglês usada na linguagem de contratos que pode ser vertida em português de duas maneiras, dependendo do contexto: “na data” e “a partir da data”. Nos exemplos (13) e (14) abaixo, podemos observar estes dois usos:

(13) *This contract is entered into* ***as of*** *June 2, 2001 by and between A and B.*

(14) *The license hereto is effective* ***as of*** *the sixth month of this Agreement*[19]*.*

Estes dois diferentes significados podem não parecer claros em uma primeira leitura, especialmente se considerarmos que as características semânticas de *enter* em (13) incluem o sentido de “tornar-se parte de”, como em <*the parties entered into an agreement*>. Entretanto, no português o verbo *enter* é geralmente traduzido por “celebrar”, um verbo que, neste contexto, não aceita como predicado um sintagma preposicional como “a partir de”. Portanto, a tradução apropriada para (13) seria (15) e não (16):

(15) Este contrato é celebrado **em** 2 de junho de 2001, por e entre A e B.

(16)* Este contrato é celebrado **a partir de** 2 de junho de 2001, por e entre A e B.

[19] Os termos *Contract* e *Agreement* serão considerados como sinônimos para fins de simplificação, visto que um contrato é o numero de termos de um (cf. Black's Law Dictionary).

Enquanto em (13) *as of* deveria ser vertido para “em”, in (14) a versão correta é aquela que considera *as of* como significando “naquela data”; portanto, deveria se vertido para o português como “a partir de”. O contexto aqui implica que há um estado de coisas que ainda não se tornou realidade, devido ao conteúdo proposicional da expressão *to be effective,* que significa “chegar a um resultado; passar a ter força de lei a partir de uma data específica”. Portanto, a versão correta para (14) deveria ser “a partir de”, como em *(*17):

(17) A presente licença *é efetiva* **a partir d**o sexto mês de vigência deste contrato.

Observa-se que enquanto *as of* está corretamente vertido em (17), o tradutor optou por usar uma referência no presente em português (é efetiva), em busca de uma equivalência sintática com o TF (14). Neste contexto específico, entretanto, a referência é no futuro, devido à expressão preposicional “a partir de”.

A fim de garantir o conteúdo proposicional, o tradutor também usou um processo de estreitamento de conceito para verter para o TA. Isto foi obtido através do acréscimo da expressão “de vigência*”*, para garantir que no TF o sentido fosse “seis meses a partir do momento que o contrato foi assinado”. Este processo de **enriquecimento** garantiu o **estreitamento** do conteúdo proposicional, graças a um acréscimo lexical no TA (cf. Carston, 1996).

Uma ilustração de como o processo de **estreitamento de conceito** também pode ser usado para desambiguação está em (18), no qual o TF é em inglês e foi traduzido por (19):

(18) X, a company **herein** represented by its Managing Quota-Holder, Mr Y.

(19) X, empresa representada **neste ato** por seu Sócio -Gerente, Sr. Y.

O termo *herein*, neste contexto, conduz à ambiguidade, pois *contract* é polissêmico: pode significar "o contrato como o documento

em si" ou "o contrato como o ato envolvendo direitos e obrigações". Portanto, em inglês (18) pode ser interpretado como em (20):

(20) X, a company **herein** ( neste contrato – o documento) represented by its Managing Quota-Holder, Mr. Y.

Ou como em (21):

(21) X, a company **herein** (neste contrato – o ato) represented by its Managing Quota-Holder, Mr. Y.

Um exemplo do processo de **alargamento de conceito** no TA (neste caso o TF está em português), pode ser observado abaixo, onde a expressão “realizou um aporte à” (23) foi traduzida por *has contributed to* (24):

(23) Antes da Data do Fechamento, o Vendedor **realizou um aporte à** Sociedade de todos os ativos fixos relacionados no Anexo 1.

(24) Prior to the Closing Date, the Seller **has contributed to** the Company all fixed assets listed in Schedule 1.

No TA, o uso de *has contributed to* carrega um significado muito mais amplo, como em <*to give (money, time, knowledge, assistance, etc.) along with others to a common supply, fund, etc., as for charitable purposes*>[20](cf. Black´s Law Dictionary). Em sentido bem mais específico, é usado na linguagem especializada de contratos de seguros como *contribution* sendo <*the method of distributing liabilities, in case of loss, among several insurers whose policies attach to the same risk*>[21]. O verbo *to realize* em direito corporativo significa <*to convert noncash assets into cash assets*>[22]; portanto, *to realize a*

---

[20] Dar (dinheiro, tempo, conhecimento, assistência, etc.) juntamente com outros para um fundo comum, para fins de caridade).
[21] O método de distribuir passivos, em caso de perdas, entre vários segurados cujas apólices também se ajustam ao mesmo risco.
[22] Converter ativos circulantes em ativos permanentes.

*supply* em inglês, possui o mesmo conteúdo proposicional de "realizar um aporte*",* em português*.* Neste caso, não ha razão para o tradutor não ter feito a tradução de forma literal, visto que havia esta possibilidade. Uma vez que o significado proposicional de *realizar um aporte* não foi transmitido no TA, o tradutor falhou em comunicar a ideia do original. O tradutor optou por **alargar o conceito**, ao transmitir somente alguns, mas não todas as implicações deriváveis do conceito codificado, conduzindo a perdas no conteúdo proposicional do TA.

No caso mencionado acima, o mesmo conteúdo proposicional poderia ter sido transmitido nas duas culturas ao se optar por uma versão direta, porém a estratégia de **alargamento** e **estreitamento de conceitos** é geralmente necessária quando há assimetrias parciais entre conceitos de diferentes culturas. Já que neste caso havia simetria, a tradução poderia ter sido orientada ao TF, como em (25):

(25) *Prior to the Closing Date, the Seller* ***realized a supply of all fixed assets*** *listed in Schedule 1 in favor of the Company.*[23]

No próximo exemplo foi feita uma tradução direta em busca de fidedignidade ao TF (26-27); entretanto, diferentemente do exemplo recém citado em (23-25), no próximo exemplo ocorreu assimetria de conceitos, o que exigiu o uso de um **conceito alargado** no TA:

(26) Todos os pagamentos estabelecidos no presente Contrato **serão efetuados** no Brasil, em moeda brasileira.

(27) All the payments set forth in this Agreement **will be effected** in Brazil, in Brazilian currency.

O **conceito alargado** em (27) não transmite o conteúdo proposicional do TF, visto que o verbo *to effect*, que significa "produzir para ter um efeito; realizar; obter; executar" não pode ser utilizado com o verbo *to pay*, como em português.

---

[23] Antes da Data do Fechamento, o Vendedor realizou um aporte à Sociedade de todos os ativos fixos relacionados no Anexo 1, em favor da Empresa.

Conforme observado, ajustes devem ser feitos a fim de garantir uma tradução fidedigna e isto pode ser obtido se lançarmos mão de estratégias tais como **alargamento** e **estreitamento de conceitos**. No entanto, antes de empregar estas estratégias, o tradutor terá de determinar o conteúdo proposicional do TF e tentar comunicá-lo no TA.

Considerando-se que um contrato é permeado de regras, pode-se supor que estas regras sejam facilmente traduzíveis, pois seguem um padrão específico. Segundo observou Sarčevic (1997), a estrutura lógica das regras jurídicas é expressa pela formula **se P, então Q**, que significa que **Q** deve ser executado somente nos casos em que as condições que constituem **P** forem preenchidas. Muitas das cláusulas de um contrato têm uma estrutura condicional, como em:

(36) Caso a Empresa X não cumpra o prazo estabelecido na Cláusula 2.3, passará então a pagar à Empresa Y 50% (cinquenta por cento) acima das quantias descritas na cláusula 2.2.

Os tradutores jurídicos são orientados a seguirem a sintaxe do TF. Quando esta não é transparente, o tradutor poderá se confundir e produzir uma oração ambígua no TA. Observemos a oração em (37), na qual o TF é em português:

(37) O saldo da Conta Gráfica apurado pelos auditores da Compradora, **caso seja positivo**, será revertido à Sociedade em 10 (dez) dias úteis contados da data de extinção da Conta Gráfica.

A tradução ficou ambígua devido ao deslocamento da oração condicional e resultou em conteúdo proposicional diferente no TA:

(38) *The positive balance of the Memorandum Account, ascertained by the auditors, **if any**, shall accrue to the Company within 10 (ten) business days as from the closing date of the Memorandum Account.*

A ambiguidade no TA em (36) reside no fato de que o tradutor deslocou o adjetivo da oração condicional para modificar o sujeito e,

portanto, teve de usar o pronome *any* para transmitir o condicional; no entanto, o pronome *any,* neste contexto, parece estar se referindo a *auditors* e não a *balance*. Sarčevic (1997) acredita que estes casos servem como alertas aos tradutores de contratos, para que estes evitem os deslocamentos de elementos na estrutura lógica do TF. Ao seguir-se sua recomendação, o texto original (37) ficaria mais claro se fosse vertido como em (39):

(39) Caso o saldo da Conta Gráfica apurado pelos auditores da Compradora seja positivo, este será revertido à Sociedade em 10 (dez) dias úteis, contados da data de extinção da Conta Gráfica.

Com base no exposto acima, a estrutura lógica de contratos internacionais deve também ser mantida, visto que este tipo de texto é um instrumento que é redigido no TF, mas irá eventualmente ser vertido para o TA. Devido ao seu objetivo principal de fazer com que as duas partes selem um acordo, neste caso cada uma das partes pertencentes a uma cultura diferente e a um sistema jurídico diferente, a situação ideal seria se a elaboração do contrato internacional fosse feita levando-se em conta tanto o TF como o TA.

O comércio exterior e a elaboração de contratos não poderão existir sem tradução jurídica e o sucesso de um empreendimento depende de um contrato bem traduzido. Por essa razão, levantei questões a fim de demonstrar que as dificuldades na tradução de contratos podem ser em parte solucionadas se forem abordadas através da pragmática e não somente através de meios convencionais de tradução. Descrevi estratégias de **alargamento** e **estreitamento de conceitos** (Carston, 1996), que estão na interface pragmático-semântica, a fim de demonstrar que há muito mais envolvido em traduções de linguagens especializadas - neste caso o texto jurídico da área de contratos internacionais entre Brasil e EUA - do que a preocupação com termos técnicos, estrutura lógica, aspectos sintáticos e semânticos.

## Referências

AUSTIN, J. 1962. How to do things with words. Cambridge, MA: Harvard University Press.

BAKER, M. (ed). 1998. Routledge Encyclopedia of Translation Studies. London: Routledge.

*BLACK'S LAW DICTIONARY* (7th ed. 2000) ed. by B. Garner. St. Paul: West Publishing Company.

CARSTON, R.; UCHIDA, S. 1998. Relevance Theory: Applications and Implications. Clevedon: Multilingual Matters.

CARSTON, R. 1996. 'Enrichment and Loosening: complementary processes in deriving the proposition expressed?' *UCL working papers in linguistics*. Vol 8: 61-88.

CODE OF PROFESSIONAL RESPONSIBILITY FOR INTERPRETERS, TRANSLITERATORS AND TRANSLATORS. 1990. Trenton, NJ: Administrative Office of the Courts, State of New Jersey.

DELISLE, P. 1988. Translation: an interpretive approach. In P. LOGAN and M. CREERY. Ottawa: Ottawa University Press.

FELLBAUM, C. Semantics Via Conceptual and Lexical Relations. In: VIEGAS, E. (ed.) *Breadth and Depth of Semantic Lexicons.* The Netherlands, Kluwer Academic Publishers, 1999.

GRICE, P. 1975. Logic and Conversation. In P. Cole & J. Morgan (eds), *Syntax and Semantics*, vol 3: Speech Acts (p. 41-58). New York: Academic Press.

______. 1981. "Presupposition and Conversational Implicature", in P. Cole (Ed.) *Radical Pragmatics*, New York: Academic Press.

GUTT, E. (1991) Translation and Relevance: Cognition and Context. Oxford: Blackwell.

LYONS, J. 1977 . Semantics. vol. 1. Cambridge: Cambridge University Press.

NEWMARK, P. 1973. 'Twenty-three restricted rules of translation'. *The Incorporated Linguist,* 12: 1.

______. 1981. Approaches to Translation. London: Pergamon.

______. 1991. About Translation. Clevedon: Multilingual Matters.

NOGUEIRA, D. 2000. Vocabulário para Direito Societário Português/Inglês. São Paulo: SBS – Special Book Services.

OBENAUS, G.1995. The legal translator as information broker. in MORRIS, M. (ed) *Translation and the Law.* Amsterdam: John Benjamins Publishing Co.

PERNA, C. 2010. Legal Translation in the light of relevance theory. In: COSTA, J.C. & RAUEN, F.J. (Org.) Topics on Relevance Theory. Porto Alegre: Edipucrs.

SAPIR, E. 1956. Culture, Language and Personality. Berkeley: University of California Press.

SARČEVIC, S. 1985. 'Translation and culture-bound terms in law'. Multilingua 4 (3): 127-33.

______. 1997. New Approach to Legal Translation. The Hague: Kluwer Law International.

______. 1998. Creativity in Legal Translation: How much is too much? In: CHESTERMAN, A, SAN SALVADOR, N & GAMBIER, Y (eds.) Translation in Context. Selected Contributions from the EST Congress, Granada 1998.

SEARLE, J. R. 1983. Intentionality: An Essay in the Philosophy of Mind. New York: Cambridge University Press. [Intentionality. Intencionalidade. Tr.: Madalena Poole da Costa. Lisboa: Relógio d'Água, 1999.]

SEQUEIROS, X. 2001. Types and degrees of interpretive Resemblance in Translation. Revista Alicantina de Estudios Ingleses. In MARTINEZ, J. AND YUS, F. (Eds) Revista Alicantina de Estudios Ingleses (14) November, 2001.Departamento de Filologia Inglesa, Universidad de Alicante.

THEODORO JR., E. 1999. 'Translating the Law: when the medium is not the message'. *ATA Chronicle*. Washington D.C.

VINAY, J. P. AND DALBERNET, J. 1958. Stilistique comparee du francais et de l'anglais. Paris: Didier.

______. 1977. Stylistique comparée du français et de l'anglais: Méthode de traduction. Paris: Didier.

WESTON, M. (1983) 'Problems and Principles in Legal Translation'. *The Incorporated Linguist.* 22(4): 207-11.

______. (1995) in Morris, M. (ed.) (1995): *Translation and the Law*, Amsterdam, John Benjamins Publishing Co.

# FAZER TERMINOLOGIA É FAZER LINGUÍSTICA[1]

*Gladis Maria de Barcellos Almeida*[2]

***Resumo***: *Há pelo menos 20 anos, a Terminologia, enquanto disciplina autônoma e como um domínio de conhecimento que integra a área da Linguística, vem se constituindo um campo de pesquisa promissor no Brasil. Entretanto, ainda não é evidente o saber linguístico que sustenta o fazer terminológico, mesmo quando as pesquisas são realizadas por linguistas, em oposição àquelas realizadas por cientistas da informação ou tradutores. Assim, este artigo tem o objetivo de apresentar, de maneira sucinta, esse saber linguístico. Discorre-se sobre algumas tarefas da atividade terminológica, de maneira a pôr em evidência o conhecimento linguístico nelas implicado. Para tanto, elegeu-se como cenário o Grupo de Estudos e Pesquisas em Terminologia (GETerm) que, desde 1999, desenvolve pesquisas nessa área. Espera-se, pois, contribuir para uma maior consolidação da Terminologia como disciplina linguística.*

Pretendemos, neste capítulo, discorrer sobre o "fazer linguístico" constitutivo da pesquisa terminológica. Para tanto, utilizaremos a experiência adquirida nos projetos encerrados e em andamento no âmbito do Grupo de Estudos e Pesquisas em Terminologia (GETerm).

Para realizar esse intento, iniciaremos por conceituar Terminologia. Em seguida, apresentaremos o GETerm para, então, percorrermos cada etapa da atividade terminológica, demonstrando toda a base linguística que está por trás de cada uma das fases do trabalho.

---

[1] A autora agradece a Ariani Di Felippo a leitura atenta e as sugestões sempre pertinentes. Eventuais impropriedades são de responsabilidade exclusiva da autora.

[2] Universidade Federal de São Carlos (UFSCar) – Centro de Educação e Ciências Humanas – Departamento de Letras e Pós-Graduação em Linguística – 13565-905 – São Carlos – SP – Brasil. Endereço eletrônico: gladis@ufscar.br

## 1. O que é Terminologia?

O termo “terminologia” pode ter duas acepções distintas. A primeira refere-se ao conjunto vocabular próprio de uma ciência, técnica, arte ou atividade profissional (TERMISUL[3]; *O Pavel*), como a terminologia da Fisioterapia, do Direito, da Música, dos Artefatos de Borracha, etc. A segunda acepção designa não só o conjunto de práticas e métodos utilizados na compilação, descrição, gestão e apresentação dos termos de uma determinada linguagem de especialidade (=terminologia enquanto atividade) (SAGER,1993), como também o conjunto de postulados teóricos necessários para dar suporte à análise de fenômenos linguísticos concernentes à comunicação especializada, incluídos aí os termos, evidentemente (=terminologia enquanto teoria).

Para efeito de clareza, emprega-se normalmente ***t**erminologia*, com inicial minúscula, para designar vocabulário ou repertório vocabular; para terminologia enquanto atividade e teoria, emprega-se ***T**erminologia*, mas com a inicial maiúscula.

O principal objetivo da Terminologia é “dar conta do funcionamento das unidades lexicais especializadas em situações comunicativas profissionais, acadêmicas ou científicas” (LORENTE, 2004), de modo que essa comunicação “se realize de forma compreensível e sem ambiguidades em ambientes mono e/ou multilíngues.” (TERMISUL)

## 2. O GETerm como cenário

O Grupo de Estudos e Pesquisas em Terminologia[4] (GETerm), formado em 1999 no Departamento de Letras da Universidade Federal de São Carlos, integra o Diretório dos Grupos de Pesquisa do CNPq. O Grupo tem como objetivos: 1) estudar conteúdos pertinentes à Terminologia; 2) desenvolver pesquisas que gerem produtos

---

[3] Grupo de Pesquisa do CNPq sediado na Universidade Federal do Rio Grande do Sul. Criado em 1991 no Instituto de Letras. Acesse o site: http://www6.ufrgs.br/termisul/terminologia_terminografia.php

[4] http://www.geterm.ufscar.br

terminológicos em língua portuguesa, tais como: glossários, dicionários, enciclopédias e assemelhados, que satisfaçam demandas reais; 3) desenvolver ferramentas computacionais e/ou ambientes que facilitem a prática terminológica.

O Grupo conta com uma equipe multidisciplinar, envolvendo linguistas, informatas e especialistas de domínio. É importante ressaltar que, desde 2002, o GETerm tem recebido valiosas colaborações de pesquisadores do Núcleo Interinstitucional de Linguística Computacional[5] (NILC), sediado no Instituto de Ciências Matemáticas e da Computação (ICMC) da USP/São Carlos. A meta do GETerm é aliar a Terminologia à Informática de forma a gerar produtos terminológicos mais confiáveis e em menor tempo.

No âmbito de GETerm, vimos tentando sistematizar um método de trabalho que, além de ser eficiente para a geração de produtos terminológicos, estivesse em consonância com os postulados de uma Terminologia de orientação descritiva, fundamentada em princípios da Linguística. Essa escolha não é sem razão, haja vista que uma teoria descritiva de base linguística parece ser muito mais adequada ao contexto brasileiro: país monolíngue com grande variedade dialetal sem outra(s) língua(s) oficial(is) minoritária(s) ou ameaçada(s) (ALMEIDA, 2006).

Filiar-se a uma teoria de orientação descritiva significa, fundamentalmente, abandonar "o tratamento prescritivo das terminologias em favor de enfoques descritivos capazes de entender o léxico especializado como um elemento natural das línguas naturais" (KRIEGER e BEVILACQUA, 2005). E fazer Terminologia fundamentada em princípios da Linguística significa, como ponto de partida, revisitar a definição de **termo**. Em razão de o termo ser o objeto por excelência da Terminologia e de que uma das tarefas mais comuns dessa disciplina é construir repertórios terminológicos, iniciaremos a exposição das etapas de trabalho pela discussão do que vem a ser termo dentro de uma perspectiva linguística.

---

[5] http://www.nilc.icmc.usp.br

## 3. Afinal, o que é *termo*?

Uma das grandes dificuldades enfrentadas pelo terminólogo que pretende desenvolver um produto terminológico[6] é, sem dúvida nenhuma, o terreno movediço que há entre **termo** (unidade dos discursos especializados) e **palavra** (unidade da língua geral). Ora, que critérios devemos utilizar para efetuar a contento a tarefa de seleção de termos em textos especializados? Dito de outro modo: como saber, ao certo, se aquela unidade selecionada é um **termo**, já que o terminólogo, na maioria das vezes, não é um especialista da área que está sendo objeto de investigação? Essa dificuldade diz respeito sobretudo: a) àqueles termos que são utilizados também na língua geral por um não-especialista, b) às lexias complexas que, muitas vezes, são apenas uma combinatória de palavras ou um sintagma discursivo, ex.: *suporte queimado, esmalte viscoso, etc.* Como saber o que realmente é termo num texto especializado, para elaborarmos o inventário terminológico acerca de determinado domínio?

Esses questionamentos nos levam a revisitar a definição de termo. Para tanto, selecionamos duas obras que têm regulamentado a pesquisa terminológica em vários países, a saber: *Terminology work - Vocabulary - Part 1: Theory and application*, ISO 1087 (2000), *Vocabulaire systématique de la terminologie* (Office de la langue française, Québec, 1985).

A primeira define termo como sendo “Designação de um conceito numa língua de especialidade por meio de uma expressão linguística.” (trad. nossa). Para a segunda, o termo é uma “unidade significante constituída de uma palavra (termo simples) ou de mais de uma palavra (termo complexo) e que designa um conceito de maneira unívoca no interior de um domínio de especialidade.” (trad. nossa)

Por essas definições, percebemos que o aspecto formal é o critério levado em conta, já que ambas se utilizam de unidades léxicas tais como e*xpressão linguística* e *unidade significante constituída de uma palavra ou de várias palavras*. Se o critério formal fosse suficiente,

---

[6] Como produto terminológico, consideramos: glossário, dicionário, base de dados terminológicos, ontologia, etc.

não haveria equívocos na identificação de termos em *corpus*[7], pois de imediato seria possível reconhecer marcas formais, principalmente no que concerne aos níveis morfológico e lexical. Essa facilidade se observa quando estamos diante de uma formação marcadamente técnico-científica, como as que utilizam morfemas greco-latinos, posto que o nível morfológico já é suficiente para indicar que se trata de um termo e não de uma palavra. Observe-se o exemplo na Fig. 1:

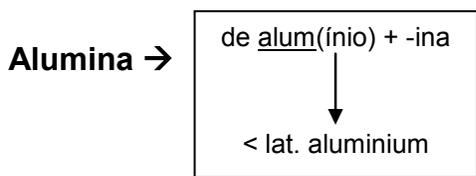



Figura 1: etimologia do termo *alumina*

Nesse exemplo, o critério formal (nível morfológico) é válido para distinguir *termo* de *palavra*. Outra possibilidade é utilizar o nível lexical e constatar os paradigmas derivacionais.

Tabela 1: Paradigmas derivacionais (adaptado de Almeida, 2000)

| P | B | VT | SN/SV |
|---|---|---|---|
| | trabalh | á | vel |
| | trabalh | a | bilidade |
| | flocul | a | nte |
| e | flocul | a | nte |
| | flocul | a | r |
| de | flocul | a | ção |
| | dispers | a | nte |
| | dispers | a | r |

[7] Neste artigo, estamos utilizando o termo corpus na acepção de Sinclair (2005): *"A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research."*

| P | B | VT | SN/SV |
|---|---|---|---|
| | esmalt | e | |
| | esmalt | a | r |
| | esmalt | a | ção |
| | viscos | i | dade |
| | viscos | í | metro |

**P** = prefixo
**B** = base
**VT** = vogal temática
**SN/SV** = sufixo nominal; sufixo verbal

Como se observa na Tab. 1, a partir da ocorrência de um dos termos, podemos inferir que as demais ocorrências constituam termos também.

Infelizmente isso não é possível com a grande maioria dos termos originários da língua geral, termos esses que não têm marcas formais para facilitar a sua recolha em textos especializados, como por exemplo: *forno, secador, peneira, biscoito,* unidades da terminologia de Revestimento Cerâmico.

Isso ilustra as dificuldades em identificar *termo* utilizando critérios estritamente formais. Que critérios devemos levar em conta para distinguir um *termo* de uma *palavra*, já que a partir de uma perspectiva linguística todos são igualmente signos da língua natural? Não existe, pois, um conjunto de termos isolados constituindo uma língua marginal à língua geral; o que há são signos da língua natural que se realizam ora como palavras, ora como termos, dependendo da temática, dos usuários, da situação comunicativa (CABRÉ, 1999; 2003). O que distingue, portanto, termo de palavra são **critérios pragmáticos**. Em outras palavras: quem diz o quê? Para quem? Em que situação? Ou como afirmou Neves (2009[8]): "Tudo aquilo que se diz, se diz em algum contexto de uso."

---

[8] Conferência de Maria Helena de Moura Neves, realizada no âmbito do Congresso de Linguística realizado em sua homenagem, em agosto de 2008 na UNESP/Araraquara, SP. Para maiores informações, conferir o sítio http://www.estudoslinguisticos.ibilce.unesp.br/eventos/congressolinguistica/index.php

Essa “redefinição” de termo permite-nos, então, constatar que a denominação “língua de especialidade” acaba sugerindo uma fragmentação e uma marginalização inadequadas, como se fosse um subsistema linguístico separado. O que temos, na verdade, é a expressão das ciências e das técnicas pela língua natural.

Se termo é assim concebido, então a sua identificação deve sempre ser feito nos contextos de uso, isso implica necessariamente a elaboração de um *corpus*, de maneira que seja possível observar os termos *in vivo* (De BESSÉ, 1997), ou nas palavras de Cabré (2004): *“La terminologia (...) debe recoger los términos originales a partir de los usos que los especialistas hacen de ellos.”* Mas o que há de linguístico por trás da tarefa de organização de um *corpus*?

## 4. A organização de *corpus* como um exercício linguístico

A elaboração de um *corpus* para pesquisas terminológicas (sobretudo naquelas cujo objetivo é a construção de dicionários, glossários, vocabulários, ontologias, bases terminológicas, etc.) exige o cumprimento de uma série de requisitos, já que é a partir do *corpus* compilado que: a) se extraem os termos e suas eventuais formas variantes; b) se observam as colocações e as fraseologias próprias de um discurso especializado; c) se infere as relações semânticas entre os termos de maneira que seja possível a elaboração de uma eventual ontologia; d) se observa o termo em todos os seus contextos de ocorrência, sendo possível inferir traços semânticos recorrentes para redigir a definição terminológica.

Uma temática linguística que é posta em evidência na seleção de textos que comporão o *corpus* diz respeito às noções de gênero discursivo – ou gênero textual, de acordo com a perspectiva de Marcuschi (2008) –, já que o cuidado nessa seleção cumpre os requisitos *amostragem* e *representatividade*, imprescindíveis para a Linguística de *Corpus* (BIBER, 1993; KENNEDY, 1998; RENOUF, 1998; BERBER SARDINHA, 2004; SINCLAIR, 2005). Segundo McEnery e Wilson (1996), um *corpus* deve ter uma amostragem suficiente da língua ou variedade de língua que se quer analisar para obter-se o máximo de representatividade desta mesma língua ou

variedade de língua. Assim, é necessário ter em conta qual noção de gênero está por trás da seleção dos textos. O pesquisador pode optar, por exemplo, pela perspectiva interacionista e dialógica de Bakhtin (1997), ou pela sistêmico-funcional de Halliday e Hasan (1989), ou a socio-retórica de Swales (1990), ou ainda a sociodiscursiva de Bronckart (1999).

Após a escolha da perspectiva teórica que fundamentará a seleção dos textos, iniciam-se as atividades de compilação, manipulação e nomeação dos arquivos de textos; pedidos de permissão de uso (quando for o caso); e anotação (se for relevante para a pesquisa) (ALUÍSIO e ALMEIDA, 2006; ALMEIDA e CORREIA, 2008).

Com o *corpus* compilado e pronto para extração de termos, podem ser iniciadas as análises morfolexicais.

## 5. Aspectos morfolexicais das terminologias

Conforme já observado em Almeida e Vale (2008), nas terminologias que têm sido objeto de estudo do GETerm, constatamos que há padrões morfológicos específicos. A terminologia de Revestimento Cerâmico, por exemplo, possui uma alta frequência de lexias complexas (*argila refratária aluminosa, análise granulométrica por peneiramento, resíduo em malha*, etc.) e lexias simples com determinados morfemas que podem servir como identificadores de termos, como os sufixos derivacionais *-agem, -ção, -mento* que podem indicar que o conceito veiculado é um processo: *secagem, moagem, britagem, ensilagem*, etc; *trituração, atomização, defloculação,* etc; *peneiramento, destorroamento, envelhecimento*, etc.

Na terminologia da Fisioterapia, o que é realmente produtivo são as formações eruditas, já que essa terminologia tem muitos termos emprestados da Medicina, ou seja, contém termos cujos morfemas, tanto radicais quanto afixos, têm origem grega ou latina, como por exemplo: *artr(i/o)-* (do gr. *árthron*) – elemento de composição que significa 'juntura'; 'articulação': *artralgia, artrectomia, artrite, artrocentese*, etc.; *-óide* (do gr. *-(o)eîdos*, pelo lat. cient. *-īdes*) – sufixo que pode significar 'aspecto ou forma de', 'semelhante a', 'relativo a': *deltóide, escafóide, articulação elipsóide, articulação trocóide*, etc.; *-ose*

(do lat. cient. *-osis* < *gr. -ōsis*) – sufixo muito produtivo na Anatomia e na Medicina, utilizado para formar os substantivos de diversos processos patológicos e doenças, agregando-se a uma base que designa quer o órgão, quer parte dele: *artrose, aterosclerose, escoliose, hiperlordose*, etc[9].

A terminologia da Nanociência e Nanotecnologia[10], por sua vez, apresenta lexias simples e complexas, entretanto, uma das características é a alta frequência do prefixo[11] *nano-*[12] (18.370 ocorrências num *corpus* de 2.565.490 palavras) em muitas unidades: *nanocompósitos, nanocristais, nanoeletromecânicos, nanossistema biológico, nanopó cristalino, filme nanoestruturado, nanotubo de carbono, nanopartícula de dióxido de titânio, imunonanoconchas*, etc.

Além de o prefixo *nano-* figurar como forma presa (unindo-se a uma base com ou sem hífen), aparece também como forma livre, neste caso, exercendo as funções de substantivo e de adjetivo. Observem-se os exemplos:

1. substantivo – nano-segregação, nanopartículas; e ocorrendo como forma livre: Estima-se também que sejam criados mais de cinco milhões de empregos em “nano” nos próximos cinco anos.
2. adjetivo – material nanocristalino, sistema nanoeletromecânico; e ocorrendo como forma livre: escala nano, ou ainda: Para 2014, recente estudo internacional aponta que o mercado de produtos “nano” será de cerca de US$ 2,3 trilhões.

---

[9] As observações etimológicas referentes aos morfemas citados no texto foram obtidas nos dicionários Ferreira (2004) e Houaiss e Villar (2001).

[10] Terminologia elaborada no âmbito do projeto “Terminologia em Língua Portuguesa da Nanociência e Nanotecnologia: Sistematização do Repertório Vocabular e Elaboração de Dicionário-Piloto – NANOTERM”, com apoio do CNPq (cf. COLETI *et al*., 2007).

[11] Foram encontrados no *corpus* alguns casos em que o formante *nano* ocorre na posição de base, tais como: *escala micro-nano, dispositivos plasmo-nano, catenano*, etc, como se poderá observar na sequência deste artigo.

[12] O prefixo *nano-* é assim definido pelo dicionário Houaiss (HOUAISS e VILLAR, 2001): “nano-: do SI, simbolizado por n, do gr. nánnos,é,on 'de excessiva pequenez' ou nânos 'anão', adotado na 11ª Conferência Geral de Pesos e Medidas, de 1960 (resolução nº 12), equivalente a um multiplicador 10-9, ou seja, milésimo milionésimo (na nomenclatura tradicional brasileira bilionésimo) da unidade indicada.”

O que observamos nessas análises de distintas terminologias é que os processos de formação de palavras mais produtivos ainda são os vernáculos, tais como: derivados prefixais, derivados sufixais, compostos e compostos sintagmáticos, confirmando as pesquisas realizadas pelo grupo TermNeo[13] (ALUÍSIO *et al.*, 2006).

Outro aspecto digno de nota é que, ao contrário de que se imagina e do que dizem os discursos midiáticos, os empréstimos do inglês não são tão preponderantes. Para se ter uma ideia, de 765 termos de Revestimento Cerâmico, apenas 9 são anglicismos, o que equivale a cerca de 1,2%: *ball-clay, china-clay, flint-clay, fire-clay, insert, creta print, euroline, mesh* e *sistema rotocolor*. Há na verdade mais termos oriundos do italiano e do francês: *muratura, bombatura, festone, mármore travertino, bordura, mate, tamização, terracotta/cotto, grês, tozzeto,* etc. Na terminologia da Fisioterapia, dos 1.200 termos que compõem o repertório, apenas 9 são anglicismos (ou 0,75%): *biofeedback, hot pack, shaking, shuttle test, spray, maitland, leg-press, splint, rash cutânea*. No vocabulário da Nanociência e Nanotecnologia, Coleti (2009) analisou 10% dos termos mais frequentes de um total de 3.069 termos. Para essa seleção, foram consideradas as listas de uni, bi, tri e tetragramas (termos compostos de 1, 2, 3 e 4 palavras respectivamente), em seguida, foi feita a seleção de 10% dos termos mais frequentes em cada uma das listas, o que correspondeu a: 80 unigramas, 59 bigramas, 60 trigramas, 16 tetragramas, perfazendo um total de 315 termos, desse total, apenas 13 termos são empréstimos ou contêm empréstimos em sua composição sintagmática: *campo de stokes, chip, conformação por spray, frequência do laser, frequência do laser escravo, frequência do laser mestre, ingap, laser, laser escravo, laser mestre, laser semicondutor, spin, stokes.* Se considerarmos que o item léxico *laser*[14] já está dicionarizado nos

[13] Acrônimo do projeto "Observatório de Neologismos Científicos e Técnicos do Português Contemporâneo", coordenado por Ieda Maria Alves, criado em 1988 e sediado no Departamento de Letras Clássicas e Vernáculas da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo (FFLCH-USP).

[14] Proveniente do inglês *laser* (1960), acrônimo de *Light Amplification by Stimulated Emission of Radiation* (amplificação de luz por emissão estimulada de radiação) (HOUAISS e VILLAR, 2001).

grandes repertórios da língua portuguesa, teríamos na verdade 6 empréstimos (todos anglicismos). Há ainda que considerar que *stokes* se repete, pois constitui um unigrama e um bigrama, portanto, temos 5 ocorrências de empréstimos nos grupos dos 10% mais frequentes, o que corresponde a 1,6%.

Assim, ao se analisar terminologias do português, desvenda-se na verdade o próprio funcionamento da língua, o que para nós não constitui uma novidade, uma vez que os termos são signos linguísticos e, portanto, estão afeitos às regras internas da língua. O que, a propósito, só vem confirmar o fato de que os repertórios terminológicos não estão separados da língua geral, mas são parte constitutiva dela:

> Parece ser que existe consenso en la idea de que la diferenciación entre el término y el no término es pragmática, puesto que los términos se forman del mismo modo que el resto de los elementos léxicos de la lengua: siguen los tipos de estructura que el sistema permite, emplean los mismos recursos de formación de palabras, y se someten a las mismas reglas de combinación y a sus restricciones. Son, en fin, tan signos lingüísticos como cualquier otro. (Castillo, 1998, p.99)

## 6. A semântica por trás das ontologias de domínio e da redação da definição terminológica

Os conceitos desempenham importante papel em qualquer projeto terminológico, já que eles não estão isolados, fazem parte de um campo especializado e relacionam-se com outros conceitos, formando uma rede ou estrutura conceitual ou ontologia. São estas ontologias — representando um conjunto sistematizado dos conceitos — que descrevem um domínio do conhecimento especializado.

Para a elaboração da ontologia, há que se conhecer o campo especializado com o qual se está trabalhando, ou então, ter assessoria permanente dos especialistas da área, pois é necessário identificar os conceitos nos textos, agrupá-los em distintos campos semânticos e estabelecer as relações entre eles.

Numa pesquisa terminológica, a ontologia é fundamental para: 1) permitir a modelagem do conhecimento num formato que possa ser reutilizado em outras pesquisas e/ou aplicações computacionais; 2) possibilitar uma abordagem mais sistemática de um domínio; 2) circunscrever a pesquisa, já que todas as ramificações do domínio são previamente consideradas; 3) delimitar o conjunto terminológico; 4) determinar a pertinência dos termos, pois separando cada grupo de termos pertencente a um determinado campo semântico, poder-se-á apontar quais termos são relevantes para o trabalho e quais não são; 5) prever os grupos de termos pertencentes ao domínio, como também os que fazem parte de matérias conexas; 6) definir as unidades terminológicas de maneira sistemática; 7) controlar a rede de remissivas (TERMCAT, 1990; ALMEIDA, 2000).

Há que se enfatizar que essa delimitação do campo é necessária porque à medida que se vai modelando o conhecimento especializado, vai-se explicitando uma determinada visão cultural e científica da realidade (CABRÉ, 1993). Isto porque a delimitação de um campo especializado é feita segundo: 1) os objetivos do trabalho terminológico; 2) o público-alvo que se quer atingir; e, sobretudo, 3) os critérios utilizados para “recortar” o conhecimento de determinada maneira. Portanto, a partir do “recorte”, tem-se uma ontologia específica.

A ontologia constitui, então, uma representação da realidade no âmbito do domínio que se toma como objeto de estudo. Essa representação procura recolher e organizar todas as ramificações que são próprias do referido domínio, de modo a refletir, em forma de esquema, a realidade da área em questão.

Assim, a modelagem do conhecimento por meio de ontologias tem-se revelado uma atividade bastante relevante e necessária em distintas áreas, uma vez que as relações semânticas contidas nas ontologias permitem aos sistemas computacionais a recuperação da informação de maneira mais eficiente.

Ao organizar esses sistemas semânticos, consideram-se basicamente dois tipos de relações paradigmáticas nos âmbitos de especialidade: as relações de abstração (gênero-espécie) e as relações ontológicas ou partitivas (parte-todo), ambas baseadas num modo de organização hierárquica. Todavia, na grande maioria dos campos

especializados, há que se considerar também outros tipos de relações, como as relações causais (causa-efeito), as relações genéticas (produtor-produto), entre outras, que não estão baseadas nos princípios de dependência hierárquica. Segundo Sager (1993), atualmente se admite que para aplicações práticas pode-se estabelecer qualquer tipo de relação conceitual e transformá-la, inclusive, num requisito para uma necessidade concreta. Por exemplo, um objeto pode relacionar-se com sua origem geográfica, sua substância material, seu método de produção, seu uso e função, etc. A configuração dos sistemas semânticos com suas complexas relações depende do objetivo a que se propõe a análise semântica, ou seja, a Terminologia não trata os sistemas como esquemas fechados e absolutos, mas como sistemas criados para o fim específico de facilitar a comunicação.

Um modelo eficiente de representação semântica pode ser, por exemplo, a Estrutura *Qualia* de Pustejovsky (1995). A Estrutura *Qualia* (= significado das palavras) como um dos níveis de representação da informação lexical dá relevo para o aspecto relacional de uma unidade léxica. Nas palavras de Pustejovsky (1995, p. 76), “de algum modo o léxico gerativo considera todos os itens lexicais como relacionais num certo grau, mas a maneira como essa característica é expressa funcionalmente será diferente, é claro, de categoria para categoria, assim como entre classes semânticas.” (tradução nossa). A Estrutura *Qualia* especifica quatro aspectos essenciais (ou papéis) que auxiliam a determinar o significado de uma palavra:

- CONSTITUTIVO: estabelece a relação entre um objeto e suas partes constituintes;
- FORMAL: distingue um dado objeto num domínio mais amplo ou mais geral;
- TÉLICO: exprime o propósito e a função do objeto;
- AGENTIVO: aponta fatores que envolvem a origem do objeto ou a causa para o objeto acontecer, ocorrer ou existir.

Segundo o autor, há dois pontos importantes que devem ser considerados nos papéis *qualia*: 1) toda a categoria expressa uma estrutura *qualia*; 2) nem todo item lexical tem um valor atribuído para cada papel *qualia*.

Observem-se, por exemplo, os termos *silicato de sódio anidro, silicato de sódio hidratado* e *silicato de sódio penta-hidratado*. Todos fazem parte da classe genérica *silicato de sódio*; entretanto, a diferença entre eles está na quantidade de água na sua fórmula. Assim, de acordo com os papéis *qualia*, teríamos:

*silicato de sódio anidro* → [FORMAL = silicato de sódio]; [CONSTITUTIVO = sem água]

*silicato de sódio hidratado* → [FORMAL = silicato de sódio]; [CONSTITUTIVO = com água]

*silicato de sódio penta-hidratado* → [FORMAL = silicato de sódio]; [CONSTITUTIVO = com 5 moléculas de água]

A atribuição de valores nos papéis *qualia* vai, portanto, identificando o objeto até o ponto em que ele possa ser distinguido dos demais e individualizado. Assim, a estrutura *qualia* provê um modelo (*template*) útil para a representação da informação lexical e aplicável computacionalmente.

Como se pode observar, há um conhecimento semântico necessário não só para a organização da ontologia como também para uma eventual aplicação computacional[15].

No que se refere à redação da definição terminológica (DT), cumpre mencionar que constitui uma das tarefas mais complexas do trabalho, já que o terminólogo precisa dominar uma multiplicidade de conhecimentos e habilidades. Primeiramente, é fundamental que o pesquisador conheça o domínio para o qual elabora o dicionário ou glossário; é necessário, também, que domine aspectos teóricos e

---

[15] Podemos citar aqui a dissertação de mestrado de Deni. Y. KASAMA, intitulada *Estruturação do Conhecimento e Relações Semânticas: uma ontologia para o domínio da Nanociência e Nanotecnologia*, defendida em 2009 na Universidade Estadual Paulista "Júlio de Mesquita Filho", campus de São José do Rio Preto (SP). Nesse trabalho, sob orientação de Claudia Zavaglia, o autor empregou a estrutura *qualia* para a elaboração de uma ontologia e sua implementação computacional no Protégé, programa disponível gratuitamente a partir da Web e que tem sido largamente utilizado para a gestão de conhecimento e edição de ontologias. O Protégé foi desenvolvido pela Stanford Center for Biomedical Informatics Research e pela Stanford University School of Medicine (http://protege.stanford.edu/).

metodológicos da Terminologia enquanto disciplina; além disso, espera-se que o terminólogo tenha noções de Linguística, posto que são acionados conhecimentos de Linguística Textual, Análise do Discurso e demais subáreas que têm o texto como objeto de estudo, afinal, a DT é, antes de tudo, um texto; por fim, o pesquisador deve conhecer a língua em que desenvolve o produto terminológico, já que a tarefa de elaboração da DT é um verdadeiro exercício de redação (ALMEIDA *et al.*, 2007).

Para além das questões textuais, um fator extremamente relevante é a seleção dos traços semânticos dos conceitos cujos termos serão definidos. Deve-se partir inicialmente da ontologia, que permite a visualização das relações de hiperonímia, hiponímia e co-hiponímia, visto que a localização do termo facilita sobremaneira a organização da definição.

Assim, para a redação de DT do tipo *gênero próximo e diferença específica*[16], selecionam-se os termos de determinado campo semântico (da ontologia) e observam-se seus traços[17] semânticos constitutivos nos contextos de uso, o que se pode facilmente recuperar a partir do *corpus*. Esse procedimento facilita a observação dos traços recorrentes, para que se possa estabelecer com segurança quais traços são imprescindíveis para a adequada descrição do conceito e em que ordem devem ocorrer no texto final da DT. Por isso, quanto maior o número de contextos para cada termo, melhor. Em princípio, isso deve ser feito com todos os termos do mesmo campo semântico, com o objetivo de atestar se os traços observados ocorrem com todos ou com a maioria dos

---

[16] Como afirma Sager (1993), deve haver vários tipos de definição de conceitos, tipos que devem ser utilizados conforme a natureza do conceito que se deve definir e conforme o fim específico da definição. Ainda segundo o autor, a posição teórica tradicional é de que o método adequado para a elaboração da DT é o modelo clássico de *gênero próximo e diferença específica* (GPDE). De fato, esse modelo tem-se mostrado bastante produtivo em nossas pesquisas no GETerm, desde que combinado com outros tipos, já que a explicitação das diferenças específicas constitui, muitas vezes, outros tipos definitórios, por exemplo *síntese*, que na tipologia de Sager (1993) equivale à descrição do conceito. Entretanto, é preciso admitir que há conceitos que exigem outros tipos definitórios, não cabendo, portanto, no modelo GPDE, como na grande área das ciências humanas (Desmet, 2002).

[17] Tais como: aspecto, origem, propriedade, finalidade, emprego, etc.

termos daquele campo. A partir do elenco dos traços, estabelece-se uma ordem de relevância e cria-se uma espécie de modelo ou gabarito que deverá orientar a redação da DT de todos ou da maioria dos termos do campo nocional escolhido (ALMEIDA *et al.*, 2007).

Nota-se, pois, que a Semântica também está subjacente a todas as fases que constituem a redação da DT.

## 7. À guisa de conclusão

Em vista de tudo o que foi assinalado até agora, cabe-nos fazer algumas considerações. Primeiro, é importante que se diga que um grupo de pesquisa que desenvolve Terminologia de base linguística não se confunde com uma “editora de dicionários técnicos”. Pesquisadores que se inscrevem nessa linha estão interessados em fenômenos linguísticos (fonológicos, morfológicos, sintáticos, semânticos, textuais, pragmáticos, discursivos, etc.) da(s) língua(s) por meio da(s) qual(is) as ciências e as técnicas veiculam seu conhecimento.

Questões de pesquisa que sempre podem ser feitas são: como os discursos especializados se apropriam da estrutura linguística para comunicar o saber científico? Que aspectos da língua são postos em evidência? Que diferenças se podem observar de domínio para domínio? Há fenômenos linguísticos mais frequentes no discurso da Medicina do que no de Computação? Que estratégias morfolexicais são empregadas para o ato da nomeação? Isso muda de domínio para domínio? Que aspectos retóricos caracterizam cada discurso especializado? Quais os *modos de dizer* de cada domínio?

Por fim, fazer Terminologia descritiva de base linguística exige o cumprimento de diversas e variadas tarefas, mas a diferença é que todas elas devem ser sustentadas por quadros teóricos provenientes da Linguística.

## 8. Referências bibliográficas

ALMEIDA, G. M. B. Teoria Comunicativa da Terminologia: uma aplicação. Araraquara, vol. I, 290 p.; vol. II, 86 p. Tese (Doutorado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras, Câmpus de Araraquara, Universidade Estadual Paulista, 2000.

ALMEIDA, G. M. B. A Teoria Comunicativa da Terminologia e a sua prática. Alfa (Araraquara), v. 50, p. 81-97, 2006. Disponível em http://www.alfa.ibilce.unesp.br/download/v50-2/06-Almeida.pdf. Acesso em setembro/2009

ALMEIDA, G. M. B.; CORREIA, M. Terminologia e corpus: relações, métodos e recursos. In: Stella E. O. Tagnin; Oto Araújo Vale. (Org.). Avanços da Linguística de Corpus no Brasil. 1 ed. Sao Paulo: Humanitas/FFLCH/USP, 2008, v. 1, p. 63-93.

ALMEIDA, G. M. B.; SOUZA, D. S. L.; PINO, D. H. P. A definição nos dicionários especializados: proposta metodológica. Debate Terminológico, v. 3, p. 1-20, 2007.

ALMEIDA, G. M. B.; VALE, O. A. Do texto ao termo: interação entre Terminologia, Morfologia e Linguística de Corpus na extração semi-automática de termos. In: Aparecida Negri Isquerdo & Maria José Finatto. (Org.). As ciências do Léxico: Lexicologia, Lexicografia e Terminologia. 1 ed. Campo Grande: Editora da UFMS, 2008, v. IV, p. 483-499.

ALUÍSIO, S. M.; ALMEIDA, G. M. B. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa linguística. Calidoscópio (UNISINOS). Vol. 4, n. 3 , p. 155-177, set/dez 2006. Disponível em: http://www.unisinos.br/publicacoes_cientificas/images/stories/pdfs_calidoscopio/vol4n3/art04_aluisio.pdf. Acesso em: setembro/2009.

ALUÍSIO, S. M.; ALVES, I. M.; ARAUJO, M.; MARONEZE, B. O.; PARDO, T. A. S. Corpora e extração automática de neologismos. In: Jornada da Rede Panlatina de Terminologia (REALITER´06), 2007, Rio de Janeiro. Anais da Jornada da Rede Panlatina de Terminologia (REALITER´06). Paris (França): União Latina, 2006. Disponível em: http://www.realiter.net/spip.php?article543. Acesso em setembro/2009.

BAKHTIN, M. Estética da criação verbal. 2ª. ed. São Paulo: Martins Fontes, 1997.

BERBER SARDINHA, T. Linguística de corpus. São Paulo, Manole, 2004.

BIBER, D. Representativeness in Corpus Design. Lit Linguist Computing, 8:243-257, 1993.

BRONCKART, J. P. Atividades de linguagem, textos e discursos. Por um interacionismo sócio-discursivo. São Paulo: Editora da PUC-SP, EDUC. 1999.

CABRÉ, M. T. La terminología en la traducción especializada. In: GONZALO GARCÍA, C.; GARCIA YEBRA, V. (eds.) Manual de documentación y terminología para La traducción especializada. Madrid: Arco/Libros. Colección: Instrumenta Bibliológica, p. 89-122.

CABRÉ, M. T. La terminología - teoria, metodología, aplicaciones (trad. castelhana de Carles Tebé). Barcelona: Editorial Antártida/Empúries, 1993.

CABRÉ, M. T. La terminología: representación y comunicación: elementos para una teoría de base comunicativa y outros artículos. Barcelona: Institut Universitari de Linguística Aplicada, 1999.

CABRÉ, M. T. Theories of terminology: their description, prescription and explanation. Terminology, v.9, n.2, p.163-200, 2003.

CASTILLO, R. A. Ideas sobre el trabajo terminográfico. In: MATEUS, M. H. & CORREIA, M. (org.) Terminologia: questões teóricas, no. 4, Lisboa: Publicações Europa-América, 1998.

COLETI, J. S. A terminologia da Nanociência e Nanotecnologia: descrição e análise morfológica. Trabalho de Conclusão de Curso. São Carlos: Curso de Letras, Universidade Federal de São Carlos, 2009.

COLETI, J. S.; MATTOS, D. F.; GENOVES JR., L. C.; CANDIDO JR., A.; DI FELIPPO, A.; A., G. M. B.; ALUÍSIO, S. M.; OLIVEIRA JR., O. N. Compilação de corpus em língua portuguesa na área de Nanociência/Nanotecnologia: problemas e soluções. Anais do VI Encontro de Linguística de Corpus. USP – São Paulo, 6 e 7 de setembro de 2007. Disponível em: http://www.nilc.icmc.usp.br/viencontro/anais.htm. Acesso em: setembro/2009.

De BESSÉ, B. 1997. "Terminological Definitions". Handbook of Terminology Management. Eds. S. E. Wrighty y G. Budin. Ámsterdam: John Benjamins. 63-74

DESMET, I. A análise do sentido em terminologia: teoria e prática da definição terminológica. TradTerm, 8, 2002, p. 169-188.

HOUAISS, A.; VILLAR, M. Dicionário eletrônico Houaiss da Língua Portuguesa. Rio de Janeiro: Objetiva, 2001

FERREIRA, A. B. H. Novo dicionário eletrônico Aurélio versão 5.0. Curitiba: Positivo, 2004.

HALLIDAY, M. A. K., HASAN, R. Language, context and text: aspects of language in a social perspective. Oxford: Oxford University Press, 1989.

KENNEDY, G. An Introduction to Corpus Linguistics. London;New York, Longman, 1998.

KRIEGER, M. G.; BEVILACQUA, C. R. A pesquisa terminológica no Brasil: uma contribuição para a consolidação da área. Debate Terminológico, no. 1, 03/2005. Disponível em http://www.riterm.net/revista/n_1/krieger.pdf. Acesso em setembro/2009

LORENTE, M. A lexicologia como ponto de encontro entre a gramática e a semântica. In: ISQUERDO, A. N.; KRIEGER, M.G. As ciências do léxico, vol. II. Campo Grande: Editora UFMS, 2004.

MARCUSCHI, L. A. Produção textual, análise de gêneros e compreensão. São Paulo: Parábola, 2008.

McENERY, T. e WILSON, A. Corpus linguistics. Edinburgh, Edinburgh University Press, 1996.

O Pavel: Curso Interativo de Terminologia. Disponível em: http://www.btb.termiumplus.gc.ca/didacticiel_tutorial/portugues/lecon1/indexe_p.html

PUSTEJOVSKY, J. The Generative Lexicon. Cambridge: MIT Press, 1995.

RENOUF, A. (ed.). Explorations in Corpus Linguistics. Amsterdam, Rodopi, 1998.

SAGER, J. C. Curso práctico sobre el procesamiento de la terminología (trad. castelhana de Laura C. Moya). Madrid: Fundación Germán Sánchez Ruipérez/Pirámide, 1993.

SINCLAIR, J. 2005. "Corpus and Text - Basic Principles" in Developing Linguistic Corpora: a Guide to Good Practice, ed. M. Wynne. Oxford: Oxbow Books: 1-16. Disponível em http://ahds.ac.uk/linguistic-corpora/. Acesso em setembro/2009.

SWALES, J. M. Genre analysis: English in academic and research settings. Cambridge: Cambridge University Press, 1990.

TERMCAT, Centre de Terminologia. Metodologia del treball terminològic. Barcelona: Generalitat de Catalunya, 1990.

# A FAMILIARIZAÇÃO SOBRE O TEMA DA TRADUÇÃO EM MEIO À FORMAÇÃO PARA O ENSINO DE LÍNGUA ESTRANGEIRA

*Heloísa Orsi Koch Delgado*[1]

> O propósito central da educação é tornar os aprendizes responsáveis pela sua própria construção de significados. Esta construção envolve pensar, sentir e fazer, e esses três aspectos devem estar integrados na construção da aprendizagem significativa, e especialmente na criação de novos conhecimentos (Novak, 2010)[2].

***Resumo**: Neste artigo, apresento parte do desenvolvimento de minha pesquisa de doutorado em andamento, que trata da inserção do tema da Tradução de linguagem especializada em meio à formação específica para o ensino de língua estrangeira. Discuto aspectos relacionados aos objetivos gerais do estudo, e procuro articulá-los à proposta de um modelo didático para uma familiarização qualificada do estudante de Licenciatura com o tema, o qual se fundamenta em pressupostos de abordagens construtivistas. Logo após, descrevo um experimento que testa uma metodologia da aprendizagem de estratégias tradutórias, dentre elas, os mapas conceituais. O estudo mostra o potencial da metodologia proposta, para que o futuro docente de língua estrangeira tenha uma compreensão inicial sobre o processo tradutório. Para a discussão da análise das traduções dos respondentes, fiz uso de conceitos sobre equivalência tradutória e textualização e sobre os princípios de livre escolha e idiomático, que parecem estar em funcionamento em toda a produção textual.*

---

[1] Doutoranda do Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul, Professora da Faculdade de Letras da PUCRS, e-mail: heloisa.delgado@pucrs.br.

[2] *The central purpose of education is to empower learners to take charge of their own meaning making. Meaning making involves thinking, feeling, and acting, and all three of these aspects must be integrated for significant new learning, and especially in new knowledge creation (Novak, 2010).* Minha tradução.

## 1. A Tradução em meio à formação para o ensino de língua estrangeira: proposta de investigação

Nas últimas décadas, tem havido um crescente interesse pelos estudos teóricos e práticos sobre Tradução no Brasil e no mundo, levando pesquisadores, tradutores e docentes a uma série de questionamentos e reflexões sobre o ensino e aprendizagem nessa área. Uma pesquisa feita em 2001 (Darin, 2006), com professores de graduação, teve como objetivo estimular a revisão crítica da prática didática e das estratégias de ensino adotadas nos cursos voltados para a formação de tradutores. Embora os resultados mostrem que a intuição, a experimentação, o uso de dinâmicas criativas, o compartilhamento local e informal de metodologias bem-sucedidas tenham contribuído para um razoável sucesso do ensino-aprendizagem da tradução, também revelaram uma falta de consenso entre os educadores sobre os parâmetros utilizados, uma centralização do processo de aprendizagem e, enfim, uma ausência de embasamento sólido na adoção de metodologias consistentes.

A pesquisa de Darin veio ao encontro das inquietações que trago comigo há mais de vinte anos, período em que me dedico ao ensino e aprendizagem da língua inglesa e à tradução de linguagens especializadas. Essas inquietações dizem respeito à formação integral (holística e específica) de alunos dos cursos de graduação em licenciatura em Letras e me levam às seguintes perguntas:

a) Se existe uma carência de embasamento sólido na adoção de metodologias consistentes nos cursos que são especificamente voltados para a formação de tradutores, como aponta Darin, será que haveria espaço para apresentar o tema da Tradução, de seus processos e produtos, em meio à formação para o ensino da língua inglesa (doravante LI)?

b) Em que medida as ementas das disciplinas dos cursos de Licenciatura em Letras, no nosso país, contemplam o ensino de habilidades e estratégias tradutórias em meio à formação para a docência?

c) O estudante de Letras de hoje, docente do amanhã, está minimamente preparado para outras tarefas que não apenas a da

docência de uma língua estrangeira, estando apto para atuar como um professor de língua estrangeira (doravante LE), que também sabe tratar sobre temas de Tradução?

Tentando ensaiar respostas para as perguntas acima citadas, realizei um levantamento inicial sobre o número de cursos de licenciatura e bacharelado em Letras (esse último, com habilitação específica em Tradução), oferecidos no Brasil, por Instituições de Ensino Superior (doravante IES). Após breve apresentação sobre esses dados, tomando por base o par de línguas inglês-português, trato do quadro no nosso Estado, o Rio Grande do Sul.

Primeiramente, em uma sondagem geral, ainda não concluída, feita através do portal e-MEC[3] (sistema eletrônico de acompanhamento dos processos que regulam a educação superior no Brasil, do Ministério da Educação e Cultura), confirmei que existem poucos cursos de Tradução no país, cerca de 10 no total. Algumas instituições que oferecem esse curso são: Universidade Federal de Brasília, Universidade Federal do Rio Grande do Sul, Pontifícia Universidade Católica do Rio de Janeiro, Universidade Federal de Minas Gerais e Pontifícia Universidade Católica de Campinas. Além desses, existem, também, cursos de Tradução ou de Letras com habilitação em Tradução em instituições como o Centro Universitário Anhanguera, a Universidade do Sagrado Coração, a Universidade Católica Unisantos, a Universidade Metodista e a Universidade Paulista, todas localizadas em São Paulo.

No que tange a cursos de Licenciatura plena em Letras (Português/Inglês) que incluem disciplinas ou atividades de ensino de Tradução, também há poucas instituições que as oferecem, embora o número de cursos de Letras, na modalidade Licenciatura em LE Inglês seja muito grande. É o caso da PUC Minas e da PUC Campinas. Na primeira, a disciplina de *Práticas de Tradução de Língua Inglesa* é oferecida no oitavo semestre do curso e enfatiza o desenvolvimento prático de *estratégias tradutórias* (grifo meu). A segunda instituição oferece duas disciplinas chamadas *Prática Autônoma de Tradução de Textos A* e *B* e *Prática Orientada de Tradução de Textos A e B* ministradas, respectivamente, no 6º e 7º semestres do curso. Nessas

---

[3] Disponível no endereço http://emec.mec.gov.br. Acessado em março de 2010.

quatro últimas disciplinas, as ementas preveem o estudo, a aplicação e o aprofundamento da teorização sobre o processo tradutório e sua aplicação prática.

Outro exemplo é a Universidade Estadual do Rio de Janeiro, a UERJ. Essa IES oferece projetos de extensão na área da Tradução através do Escritório Modelo de Tradução Ana Cristina César. Esse projeto[4] trabalha em perspectiva amplamente universitária, desenvolvendo atividades de Ensino, Pesquisa e Extensão e vem abrindo frentes, até então não trilhadas, para docentes, pesquisadores e alunos dos cursos de Inglês, Espanhol, Francês, Italiano e Alemão do Instituto de Letras, no que diz respeito aos Estudos de Tradução. Entretanto, cabe dizer, a UERJ não oferece uma graduação específica para formar tradutores.

Falemos, agora, sobre o número de cursos de Licenciatura em Letras no RS. Tomando como referência novamente o e-MEC, foram selecionadas 24 IES que oferecem o curso de Licenciatura em Inglês ou em Português-Inglês, em diferentes localidades do nosso Estado.

Constatei que, dos 24 cursos de Licenciatura em Letras (sem habilitação em Tradução), apenas um, oferecido pela UNIFRA, em Santa Maria- RS, contempla estudos teóricos e práticos da Tradução no 5º e 6º semestres.

A PUCRS[5], a UFSM e a UNISINOS estão entre as IES que oferecem disciplinas sobre a prática de leitura e de expressão escrita. A primeira oferece *Leitura e Produção Textual* que enfatiza *a prática de leitura e produção de textos descritivos e/ou narrativos com ênfase em seus aspectos semânticos, sintáticos, pragmáticos e discursivos característicos.* A UFSM, através da disciplina de *Leitura em Língua Inglesa,* tem como objetivo *ler extensivamente textos em língua estrangeira, utilizando competências sistêmicas, textuais, estratégicas e discursivas e a UNISINOS* oferece a disciplina *O Leitor e o Texto em Língua Inglesa* que enfatiza aspectos diversos de compreensão leitora em língua estrangeira.

---

[4] Sob a coordenação de Maria Aparecida Salgueiro.

[5] A PUCRS, no seu novo currículo de 2011, incluirá na disciplina *Língua Inglesa VIII*, noções sobre estratégias tradutórias.

A ênfase maior dada nas ementas das disciplinas oferecidas pelos 24 cursos investigados é a do ensino e da aprendizagem da língua e da literatura. Os três cursos que oferecem disciplinas de leitura e produção, como as citadas no parágrafo anterior, não preveem sequer uma familiarização com conteúdos que tratem do tema da Tradução, tampouco chegam a explicitar que existe outra profissão ou ocupação relativamente próxima aos temas dos estudos da LE.

A minha preocupação central reside no fato confirmado: não há um espaço, em meio à formação docente de LE, onde os estudantes possam ter algumas noções sobre a natureza da Tradução como atividade profissional, diferenciando-a da atuação no ensino. Há, assim, apenas uma lacuna ou silêncio que, a meu ver, pode levar à percepção errônea de que essa atuação não existe ou a de que seja algo de simples execução, inerente ao aprendizado da língua.

Apesar desse quadro, penso que o contato do futuro professor de LE com o tema, em especial com o tópico de estratégias tradutórias, poderia ser benéfico à medida que o levasse a utilizá-las em prol de sua própria construção do conhecimento. Enfim, a lacuna existente, ainda que possa ser pensada como uma salvaguarda de domínios, visto que formar tradutores e formar professores são processos diferentes, poderia gerar resultados positivos ao ser preenchida com o devido critério.

Há, sim, espaços importantes que contemplam a formação do discente em Letras em várias outras áreas do conhecimento (literatura, cultura, etc.) e dentro de sua área de atuação, que é o ensino da língua inglesa (ensino e aprendizagem, metodologia, etc.), mas ao que parece, há pouco (quase nenhum) interesse na inclusão de aspectos relativos à aprendizagem de estratégias tradutórias. E, aqui caberia mais uma pergunta: se os cursos de formação específica para tradutores são tão poucos, não seria o caso dos cursos de Licenciatura tratarem do tema nem que fosse para auxiliarem a promover os cursos de Tradução como uma segunda possibilidade de formação profissional em Letras?

A pesquisa feita por Darin, referida anteriormente, ressalta que o desenho curricular dos cursos de graduação em Tradução deveria se preocupar com questões que vão além do conteúdo

programático que, normalmente, respondem a questões do tipo *o quê, como,* e *para quê.* Dentro de uma visão menos simplista, ela sugere a inserção do papel da argumentação teórico-didática no contexto universitário (*o porquê*):

> se concebemos o curso universitário como espaço por excelência destinado ao crescimento intelectual e desenvolvimento do olhar crítico, é essencial que o estudante tenha a oportunidade de refletir sobre temas e procedimentos relativos ao papel do tradutor sob a ótica de inúmeras questões da contemporaneidade. São questões como essas que podem orientar o desenvolvimento da capacidade reflexiva e crítica tão necessária para que os futuros profissionais produzam alguma mudança nas formas como nossa sociedade compreende a tradução e a identidade do tradutor. (Darin, 2006, p. 119).

Com base nas reflexões da autora e adequando-as à proposta de minha pesquisa de doutoramento, formulei a seguinte situação-problema:

*Considerando a lacuna de formação existente relativa ao tema /natureza da Tradução em meio à formação do docente de LE, e que contribuir para diminuí-la é algo que pode ser positivo, como se pode introduzir esse tipo de conhecimento para o acadêmico de licenciatura em Letras?*

À luz desse questionamento, apresento aqui uma metodologia didática da tradução de linguagens especializadas em meio à licenciatura em LE. Na direção de reconhecer e até de propor modos de fazer essa inserção é que se move a pesquisa que realizo, ainda que, para tanto, tome por base apenas uma dada situação de aprendizagem, a que envolve o contato com o texto científico entre alunos do curso superior de Licenciatura em Letras. Essa metodologia será avaliada ao final deste texto e perspectivas serão colocadas para o desenvolvimento da pesquisa. Passo agora à apresentação das origens para a proposta.

## 2. Antecedentes da proposta metodológica

O interesse pelo tema da inter-relação entre formação docente e Tradução vem apoiado numa trajetória de 23 anos como professora de língua inglesa e tradutora. Esse percurso foi enriquecido pelo trabalho como professora de língua geral e ESP (*English for Specific Purposes* – Inglês para Fins Específicos) em instituições de ensino e em empresas, além da experiência como tradutora de textos científicos e técnicos.

A experiência como professora de ESP para profissionais de Medicina e de Aviação, cujo objetivo principal era o de adquirir a proficiência leitora de artigos científicos e de manuais de manutenção e operação, fez com que a estreita relação entre o ensino de uma língua estrangeira e o ensino de estratégias de compreensão leitora desse espaço para o tema da Tradução no contexto de sala de aula.

Desde o término do meu curso de graduação (Bacharelado em Letras - Tradução – Português/Inglês/Francês) e o início da minha vida profissional e, principalmente após a minha dissertação, que foi na área da Educação, muito tenho refletido sobre as minhas próprias ações pedagógicas. Sigo uma concepção de educação baseada na pedagogia dialética, construtivista em que o ensino deve estar centrado na relação professor-aluno e no conhecimento que é gerado a partir dessa relação. Vejo o papel do professor como um mediador, agente reflexivo, que toma decisões e é criativo.

Essa visão sobre o olhar educativo sempre me causou inquietações e me conduziu a uma prática mais flexível em sala de aula, através da inserção de temas interdisciplinares e transversais, que contemplassem outros aspectos da aprendizagem e do ensino de uma língua. Um desses temas diz respeito ao papel do professor, conhecedor de uma língua estrangeira, frente à tarefa da tradução, cada vez mais emergente no mercado de trabalho desse profissional.

Levando em consideração esses antecedentes, propus os seguintes objetivos na minha pesquisa, que intenta ratificar a crença no papel positivo da inter-relação entre docência e Tradução:

- Apresentar uma metodologia de aprendizagem de estratégias tradutórias da linguagem científica, tendo em vista o aluno de licenciatura em língua inglesa.

- Reconhecer e explorar as características dessa linguagem, apresentadas no âmbito do Transtorno do Humor Bipolar (doravante THB).
- Propor a elaboração de mapas conceituais como facilitadores da aprendizagem desse campo temático e da tradução de textos científicos que o veiculam.
- Tratar do texto traduzido como textualização, tomando como base o conceito proposto por Costa (2005).
- Analisar os textos traduzidos dos respondentes, a partir do princípio idiomático e o do princípio da livre escolha de Sinclair (1991).
- Pensar em uma proposta de ambiente de aprendizagem virtual que contemple aspectos de familiarização e aprendizagem de estratégias de tradução de linguagens especializadas, para estudantes de Letras.
- Dar início à elaboração de um livro, cuja abordagem esteja centrada em estratégias ou ações que conduzam à resolução, de forma satisfatória, de instâncias problemáticas tradutórias, tendo em vista a formação do docente em LE.

## 3. Textos e linguagens científicas

A missão principal do texto científico é a de informar e isso o distingue de outros tipos de mensagens. O discurso científico, segundo Rodilla (1998), possui algumas características principais tais como:

a) **Precisão**: constitui, sem dúvida, a característica mais importante da linguagem científica. Está relacionada, em grande parte, com a precisão dos termos empregados para a elaboração da mensagem e que pode ser perdida quanto maior for a sinonímia, a polissemia e a homonímia contidas nestes termos. Esta precisão significa que o valor do termo utilizado não depende de nenhum dos fatores que possam condicionar o ato comunicativo e, graças a ela, se evita a ambiguidade e as falsas interpretações. Por exemplo, o CID 10[6] tem validade internacional em qualquer língua para evitar que

[6] Classificação Internacional de Doenças.

diagnósticos errôneos sejam feitos. Os diagnósticos são feitos a partir de sintomas patognomônicos (sintomas próprios de cada moléstia e cuja identificação permite um diagnóstico certo) com validade internacional como, por exemplo, para o diagnóstico de *surto maníaco* é necessária a presença do sintoma *euforia*, ou seja, *humor exaltado.*

O conceito de precisão da autora parece estar relacionado a um texto onde a polissemia, a ambiguidade não existem. Sabemos, no entanto, que esses aspectos ocorrem em diversos textos científicos, principalmente aqueles destinados à divulgação para leigos. Um exemplo disso são as informações encontradas no sitio *Stabilitas*[7], utilizado nesta pesquisa como um procedimento metodológico (explicação na parte da metodologia). Também há de se levar em consideração o estilo do autor, mesmo que escreva textos científicos com alta densidade terminológica e que apresentem um maior nível de precisão semântica.

b) **Concisão ou economia**: é outra característica da mensagem científica em que se usa o menor número de palavras necessárias para expressar a ideia que se deseja comunicar. Logicamente, esta propriedade não pode ir contra uma transmissão clara e eficaz do conteúdo do conhecimento. Trata-se, por um lado, de uma economia semântica relacionada com o número de palavras empregadas, cujo expoente máximo é a substituição de frases inteiras por uma única palavra através da invenção de um novo termo (ex.: hipomaníaco: surto maníaco de menor intensidade no transtorno de humor bipolar). A economia linguística ressalta as condições implícitas particulares da comunicação técnica, pois sua aplicação supõe uma série de premissas como, por exemplo, de que os participantes nesta comunicação tenham os mesmos conhecimentos do assunto. No entanto, as frases da linguagem científica podem ser longas, mais longas do que outros tipos de linguagem, devido à complexidade dos conceitos científicos, que exige um número maior de palavras para que possam ser explicados. Tomamos como exemplo o conceito de *depressão* que se caracteriza por *mais de duas semanas de humor*

---

[7] Disponível no endereço eletrônico http://www.stabilitas.kit.net. Acessado em janeiro de 2010.

*irritado ou profundamente triste, sensação de vazio, desesperança, medo, desânimo e insegurança.*

c) **Peculiaridades do vocabulário**: para muitos autores, a diferença entre a linguagem comum e a de especialidade é o léxico, enquanto que para outros, essas diferenças atingem todos os planos da língua: morfológico, sintático ou fonético. No que todos concordam, no entanto, e que é o elemento caracterizador da linguagem científica, é o vocabulário que emprega[8].

Enquanto alguns estudiosos acreditam que o que predomina no léxico científico são o verbo e o substantivo (*identificar a doença*), pois o verbo desempenharia um papel de conector modal entre o sujeito e o objeto; outros, creem que é o substantivo e o adjetivo (*humor eufórico*), os que alcançam uma frequência mais elevada de uso.

Há, também, as colocações verbais (substantivos que co-ocorrem naturalmente com determinados verbos) que se usam na linguagem científica como, por exemplo, *aderir ao tratamento* e *transmitir através da genética.* Nesses exemplos, os verbos cumprem quase que exclusivamente funções sintáticas, pois a informação semântica principal está nos substantivos.

Grande quantidade dos termos técnico-científicos se constrói a partir de prefixos gregos ou latinos, o que beneficia a comunidade internacional, acima das fronteiras linguísticas nacionais ou regionais (*maladjustment*, *premorbid).* Existe, além disso, outro procedimento de formação de termos científicos, não exclusivo, mas muito típico, que está relacionado com as diminuições (siglas, abreviações: THB – Transtorno do Humor Bipolar) e com a hiponímia (Tipo 1, Tipo 2 – tipos de THB).

Outra diferença entre o léxico científico e o comum é a sua velocidade de crescimento. Tomamos como exemplo, a antiga denominação *psicose maníaco depressiva,* para o atual *Transtorno do Humor Bipolar*. Com o aumento do conhecimento, hoje, se sabe que existem dois tipos de transtorno de humor. Novas pesquisas já ampliam este conceito para *espectro bipolar*, incluindo vários estágios da doença.

---

[8] Tanto é assim, que é possível a comunicação, em língua inglesa, que veicula o conhecimento científico, por profissionais que tenham apenas conhecimentos básicos da gramática desta língua, o que permite entenderem e serem entendidos em congressos profissionais internacionais.

No mundo da ciência e da técnica, as mudanças podem ser espetaculares, inclusive de um dia para o outro. Dessa forma, partindo de um patrimônio léxico tradicional, a linguagem comum cresce num ritmo lento, comparado ao ritmo de crescimento do vocabulário científico. Por outro lado, atualmente, a principal fonte de léxico para a linguagem comum provém do mundo da ciência e da técnica, democratizado através dos meios de comunicação.

## 4. Mapas conceituais e competência tradutória

Com o objetivo de contextualizar um estudo piloto que realizei, tomo como ponto de partida algumas noções de aprendizagem significativa, propostas por Ausubel (1968, 2000) e Cabré (1999) e de competência tradutória, por Hurtado Albir (2005).

O conceito proposto por Ausubel, em sua Teoria da Aprendizagem de Assimilação, se fundamenta na Psicologia Educacional, proposta pelo mesmo autor (1968). Ele aponta que aprendemos novos conhecimentos ao relacioná-los a uma rede de conceitos que já conhecemos, ou seja, aprender significativamente é estar ciente da natureza das relações que definimos entre objetos ou fatos, ao inseri-los na base de conhecimento já existente. No entanto, o aprendiz é que deve procurar uma maneira de integrar essas informações com aquelas já existentes em sua estrutura cognitiva. O papel fundamental do professor, em sua teoria, é o de encorajar seus alunos a isso, fazendo uso de diversas estratégias educacionais, dentre elas, os mapas conceituais, estratégia que utilizei no experimento descrito na seção *Desenho inicial da proposta para a inserção do tema da tradução em cursos de docência em LE*. A utilização de mapas conceituais, para minha pesquisa, conforme acredito, pode ser um recurso consistente para facilitar o gerenciamento de informações escritas nas linguagens especializadas, auxiliando os usuários dessa linguagem que, neste caso, são alunos de Licenciatura em Letras, a distinguir o essencial do dispensável e a adquirir habilidades determinantes para a compreensão da área temática a ser traduzida.

Nesse sentido, Cabré (1999) ressalta que a organização da linguagem utilizada para produzir determinado conhecimento por meio

de mapas conceituais é um recurso eficiente para garantir as bases para a organização de campos temático-funcionais. A estudiosa ainda salienta que os conceitos de uma mesma área especializada, mantêm, entre si, diferentes tipos de relações e todas essas relações juntas formam o mapa conceitual de determinado campo. Para isso, entende-se ser necessária a criação de um “macroesquema” que relacione os conceitos de uma área do conhecimento e funcione como “esqueleto” dos conteúdos apresentados nos textos científicos a serem traduzidos pelo especialista da linguagem.

O mapa conceitual, portanto, é uma estrutura esquemática usada para representar um conjunto de conceitos imersos numa rede de proposições. Ele é considerado como um estruturador do conhecimento, na medida em que permite mostrar como o conhecimento sobre determinado assunto está organizado na estrutura cognitiva de seu autor que, assim, pode visualizar e analisar sua profundidade e extensão. Ele pode ser entendido como uma representação visual utilizada para partilhar significados, pois explicita como o autor entende as relações entre os conceitos enunciados.

Assim, ao dispor sob a forma gráfica de um mapa conceitual, com conceitos conhecidos de uma dada especialidade, relacionando-os a outros também já conhecidos e estabelecendo uma hierarquia e/ou determinando propriedades, o aprendiz poderá organizar o seu conhecimento de maneira autônoma, retificando seu próprio raciocínio em função da construção do mapa. Para a realização da aprendizagem significativa, mapas conceituais demonstraram ser uma ferramenta adequada, porque possibilitam ao aluno (e ao professor também) desenvolver um processo cognitivo de aprendizagem, em que ele próprio orienta a aquisição de novas informações, pois elas estarão diretamente relacionadas com a estrutura de conhecimento prévio, aquela que o indivíduo possui, no momento da aprendizagem, como produto da sua integração cultural. Portanto, os mapas conceituais revelam como cada um dos aprendizes estrutura seu conhecimento em relação ao conteúdo de estudo. Existem inúmeras possibilidades de hierarquizar e relacionar os conceitos sob foco e, por isso, dois mapas diferentes sobre o mesmo assunto, podem estar igualmente corretos. Essas representações são consideradas mais que um artifício pedagógico, pois reproduzem, de

forma direta, a organização de conceitos na própria memória humana. Como exemplo disso, no final deste artigo, apresento os mapas construídos por dois respondentes deste estudo piloto, que tiveram como base os mesmos textos sobre THB.

A outra aposta teórica que tomo como referência para o meu estudo é a de aquisição de competência tradutória (Albir, 2005), quando esta ocorre de uma maneira guiada no contexto do sistema educacional: a didática da tradução. Para meu estudo, dada a peculiaridade do cenário de formação envolvido, gostaria de me referir a esta *aquisição* da competência tradutória como *a aprendizagem e familiarização* de estratégias tradutórias. A ideia é que os futuros docentes da LE construam um caminho paralelo ao do ensino e que tenham subsídios para compreender a natureza das tarefas de Tradução com propriedade e, quem sabe, se especializar nessa área tão importante nos dias atuais.

A autora salienta que qualquer abordagem educacional deve partir necessariamente de dois grandes blocos fundamentais: um enfoque no conteúdo a ser ensinado e um enfoque sobre a aprendizagem desse conteúdo. Para sustentar a base da didática da tradução a que se propõe, ela descreve um arcabouço teórico duplo: o tradutológico e o pedagógico, brevemente descritos a seguir.

No que diz respeito ao quadro tradutológico, Albir entende que qualquer abordagem formativa de disciplinas relacionadas com a tradução sustenta-se em um determinado enfoque sobre a tradução (a prática comunicativa que se deseja ensinar), a competência tradutória (os conhecimentos e habilidades requeridos para realizar esta prática) e a aquisição da competência tradutória (como esta competência é adquirida).

Para poder estudar sobre a competência tradutória em seu conjunto e dar conta da complexidade intrínseca da tradução, propôs um modelo de investigação empírico-experimental sobre a competência tradutória e sua aquisição (Hurtado Albir, 2001:394-401), revisado em 2003. Considera que a competência tradutória é um conhecimento especializado que consiste em um sistema subjacente de conhecimentos declarativos e operacionais, necessários para saber traduzir, que está composto de cinco subcompetências, a

saber: i) a bilíngue, integrada por conhecimentos pragmáticos, sociolinguísticos, textuais e léxico-gramaticais; ii) a extralinguística, composta por conhecimentos sobre o mundo em geral e de âmbitos particulares; iii) a de conhecimentos sobre tradução, composta por conhecimentos sobre os princípios que regem a tradução e sobre aspectos profissionais; iv) a instrumental, que consiste em conhecimentos operacionais, relacionados ao uso das fontes de documentação e das tecnologias de informação e comunicação (TIC) aplicadas à tradução e a v) a estratégica, que controla o processo tradutório e serve para avaliar o processo e os resultados parciais obtidos em função do objetivo final, compensar deficiências entre as subcompetências, identificar problemas e aplicar os procedimentos para a sua resolução.

Naturalmente, como trato dessas competências em meio a uma formação de docência em LE, será preciso relativizar essas indicações.

No que se refere ao quadro teórico pedagógico, a autora sugere o *enfoque por tarefas de tradução,* cujos aspectos principais são:

- Instrumentaliza o estudante, ao introduzir tarefas facilitadoras (pedagógicas) que o ajudam a resolver a tarefa final (a tradução de um determinado gênero textual, por exemplo, que, neste caso, foi a de um artigo científico), obtendo-se, assim, uma pedagogia centrada no acompanhamento de processos (os mapas conceituais elaborados pelos respondentes desta pesquisa foram avaliados pela especialista e por mim e reportados aos mesmos).
- Reforça, constantemente, através das tarefas, uma metodologia viva na qual o aluno não apenas aprende *fazendo,* e capta princípios, como também aprende a resolver problemas e adquire estratégias (as tarefas facilitadoras sugeridas, no estudo piloto, têm um caráter terminológico e serviram para que os respondentes, de forma autônoma, buscassem por estratégias de aprendizagem sobre o THB).

Revisadas, aqui brevemente, as noções teóricas envolvidas, passo agora à apresentação de minha proposta.

## 5. Desenho inicial da proposta para inserção do tema da tradução em cursos de docência em LE

A minha proposta, aqui preliminarmente relatada, está amparada em um estudo piloto que venho desenvolvendo com alunos de um curso de Licenciatura em Letras com habilitação em língua inglesa. O experimento também contou com a participação de uma médica psiquiatra, especialista em THB[9] e da minha, como tradutora. Os resumos, em língua inglesa, de artigos científicos, feitos pelas duas últimas respondentes, foram utilizados como padrão de referência para fins de análise.

O estudo consistiu em uma série de procedimentos que tiveram como objetivo a familiarização do estudante com uma série de elementos ou questões:

1) A importância de se conhecer o tema a ser traduzido *a priori*, partindo de conceitos mais amplos e de fácil compreensão, para conceitos mais específicos e complexos.
2) A natureza e o papel dos mapas conceituais, para auxiliar no entendimento do campo temático a ser traduzido e na compreensão da leitura dos textos que o veiculam.
3) A busca autônoma por subsídios que auxiliem na tarefa da tradução, ou seja, a captação de estratégias de aprendizagem.
4) A natureza do texto científico.
5) A familiarização com o tema de estratégias tradutórias.

A seguir, apresento uma descrição do experimento, começando pelas características gerais dos sujeitos da pesquisa e pelos instrumentos e procedimentos para a coleta de dados e, em seguida, pelo *corpus* utilizado.

---

[9] Dra. Carmen Vernetti, cuja *expertise* e prática dela advindas foram imprescindíveis no momento da revisão dos mapas conceituais feitos pelos respondentes e na análise das suas textualizações.

### *5.1 Os sujeitos da pesquisa*

Contei com a colaboração de dez alunos respondentes, acadêmicos do curso de graduação e pós-graduação em Letras. Possuem níveis diferentes de proficiência da língua inglesa e de experiência profissional. Durante o período da coleta dos dados (traduções de resumos de artigos científicos), os estudantes estavam cursando o sexto semestre do curso*.* Como supracitado, fazem também parte do estudo, a especialista e eu, a tradutora.

Para fins de descrição dos resultados do estudo piloto, foram selecionadas as traduções de quatro alunos respondentes (R1, R2, R3 e R4), cujos perfis serão descritos brevemente a seguir, conforme respostas fornecidas através de um questionário, da especialista e da tradutora.

| | Sexo | Formação Acadêmica | Experiência Profissional |
|---|---|---|---|
| **R1** | Feminino | Acadêmica de Letras/Hab. Português-Inglês (7º semestre) | Trabalha como bolsista do Programa de Ensino Tutorial/Letras. Não possui experiência em tradução. |
| **R2** | Feminino | Acadêmica de Letras/Hab. Português-Inglês (7º semestre) | Trabalha como técnica em uma empresa e como tradutora voluntária para um *site* que divulga notícias internacionais. |
| **R3** | Feminino | Acadêmica de Letras/Hab. Português-Inglês (7º semestre) | Trabalha como bolsista de Iniciação Científica. Não possui experiência em tradução. |
| **R4** | Feminino | Acadêmica de Letras/Hab. Português-Inglês (7º semestre) | Trabalha na Editora Artmed como revisora de ementas. Não possui experiência em tradução. |

## 6. Instrumentos e procedimentos para a coleta de dados

A coleta de dados foi conduzida de acordo com procedimentos metodológicos propostos por mim, inspirados em uma didática com ênfase na aprendizagem autodidata sobre estratégias tradutórias e baseada em tarefas, relacionada aos trabalhos de Albir (2005). Essa autora propõe um processo de formação de natureza cognitivo-construtivista centrado na aprendizagem por tarefas e cujos objetivos levem em consideração três aspectos: a competência que se pretende que o estudante possa vir a ter, o desenvolvimento de estratégias de aprendizagem e a aquisição de atitudes e valores.

Posso, portanto, à luz dessas indicações, pressupor que é viável e necessário haver o tratamento da Tradução em meio ao ensino de língua estrangeira em cursos de Licenciatura em Letras. A hipótese a ser investigada é a de que, através de um modelo didático, centrado no aprendiz e baseado em tarefas, pode-se obter uma compreensão sobre a aprendizagem de estratégias tradutórias. Essa compreensão, conforme acredito, poderá ser revelada na produção tradutória desses estudantes.

A metodologia de ensino que adotei teve as seguintes etapas:

a) criação de uma *wiki* (um conjunto de páginas da web) para os dez respondentes, a especialista, a pesquisadora[10] e a orientadora da pesquisa. A intenção de criar este espaço virtual foi a de padronizar as informações, servir de um repositório de arquivos, material de consulta e links para as tarefas a serem feitas e as já concluídas, e de um canal de comunicação entre o grupo. O endereço da *wiki* é http://pucrsteaching.pbworks.com/.

b) **solicitação de tarefas para ambos os grupos:**

b1) leitura de um artigo[11] sobre a construção de mapas conceituais, escrito por Romero Tavares (2007). O objetivo dessa tarefa foi o de apresentar os diferentes tipos de mapas conceituais existentes e suas finalidades dentro de contextos diversos.

---

[10] Eu, no papel de pesquisadora e tradutora.

[11] Texto disponível na *web* (referência no final do artigo) ou no ambiente da *wiki,* através dos links *Grupo A* e *Construindo Mapas Conceituais.*

b2) acesso ao sítio *Stabilitas*[12] para terem as primeiras noções sobre a área desconhecida, envolvida nos textos a traduzir e da linguagem empregada nesse âmbito (nível baixo de densidade terminológica). Importante salientar que a língua utilizada no material disponibilizado sobre o THB é a do português brasileiro, pois é nessa língua que os respondentes farão as textualizações.

b3) leitura de um folder que apresenta as principais ideias do THB, através de uma linguagem acessível, fornecido pela especialista (nível baixo de densidade terminológica).

b4) leitura de material adicional[13] sobre a área, escrito pelo Dr. Diogo Lara[14] para estudantes e profissionais, além do acesso ao seu sítio http://www.bipolaridade.com.br. O objetivo da tarefa foi ampliar o conhecimento sobre o THB através de textos que apresentem um nível médio de densidade terminológica.

As tarefas acima foram necessárias para que o grupo tivesse uma ideia sobre a pesquisa como um todo.

c) **divisão do grupo em dois (três em cada)**, para investigar a utilidade da metodologia por mapas conceituais, denominados *Grupo A* (que construíram os mapas e doravante denominados Grupo A/CM – com mapa) e *Grupo B* (que não construíram os mapas e doravante denominados B/SM- sem mapa). Para dar ao ambiente (*wiki*) um olhar mais organizado e ágil, foram inseridos dois *links*: um para o grupo A e o outro para o B. Dessa forma, os sujeitos acessariam apenas à página indicada aos seus grupos. Além disso, para cada página (grupo A e B), foram inseridos *links* com os seus nomes para que pudessem usufruir de uma página individual para referência e postagem dos dados.

---

[12] Esta associação não possui fins lucrativos e é formada por portadores do Transtorno do Humor Bipolar, familiares e amigos e surgiu a partir de reuniões realizadas no Hospital de Clínicas de Porto Alegre. Tem por objetivo principal a junção de forças em prol da melhoria da qualidade de vida dos portadores desta enfermidade.

[13] *Depressão e ansiedade resistente: será bipolaridade?* e Bipolaridade, Transtorno de Personalidade, *Borderline e Transtorno de Déficit de Atenção e Hiperatividade: semelhanças e diferenças.* Textos da revista Bipolaridade, do Dr. Diogo Lara, escrita exclusivamente para a classe médica. Material disponível na *web* e na *wiki.*

[14] O Dr. Diogo Lara é uma das maiores autoridades em THB do Brasil.

d) **instrumentalização do Grupo A/CM**

d1) construção de um mapa conceitual sobre o THB, utilizando os conceitos e termos que os sujeitos encontraram nas leituras feitas[15].

d2) avaliação do mapa pela especialista e postada na *wiki* com comentários sobre cada um dos mapas construídos.

d3) reformulação dos mapas já iniciados.

d4) avaliação dos mapas reformulados pela especialista e postada na *wiki* com comentários.

e) **solicitação de tarefas para ambos os grupos:**

e1) solicitação da tradução, de inglês para o português, de três resumos de artigo - foi solicitado o prazo de cerca de um mês para a conclusão dessa tarefa. Os respondentes trabalharam em suas residências e puderam utilizar os recursos que desejassem: dicionários impressos e/ou eletrônicos, gramáticas, os mapas (no caso, o Grupo A, apenas) que construíram e material de referência sobre a doença (impresso e eletrônico), fornecidos pela pesquisadora e pela especialista. Esta também se utilizou da *wiki* como canal de comunicação, fez as traduções em sua residência e teve prazo de um mês para a realização da tarefa.

f) **solicitação, a posteriori, da leitura do texto *Processamento Cognitivo "Teoria da Mente" no Transtorno Bipolar*[16]*,*** para aprofundar o conhecimento sobre a doença, e da ampliação dos mapas conceituais. Após, solicitação da tradução de um texto que possui um nível alto de densidade terminológica, retirado da Revista *Bipolar Disorder,* especializada no assunto e internacionalmente reconhecida.

## O texto de partida

Os textos de partida que utilizei para a minha pesquisa apresentam um grau de dificuldade crescente (partindo de um texto

[15] Exemplos dos mapas construídos pelos sujeitos, para esta pesquisa, no fim do artigo.

[16] Texto escrito por Hélio Tonelli, do Instituto de Psiquiatria do Paraná, publicado na Revista Brasileira de Psiquiatria, vol. 31, São Paulo, em dezembro de 2009.

com baixa densidade terminológica até um nível alto de densidade terminológica). Neste artigo, apresentarei a análise da tradução de um *abstract* (ou resumo) sobre o THB, com um nível médio de densidade terminológica, considerando que a revista, utilizada nesta etapa, não é especializada apenas nessa doença.

O texto de partida (no final deste artigo) foi escrito em língua inglesa e portuguesa, e publicado em 2004, na revista eletrônica *Revista Brasileira de Psiquiatria,* reconhecida pelo seu excelente nível científico. O texto (em inglês) tem 139 palavras (juntamente com o título) e é intitulado *The multidisciplinary team approach to the treatment of bipolar disorder: an overview.* Trata-se de um texto associado ao gênero resumo, escrito por pesquisadores brasileiros. Embora o texto date de 2004, sua importância em termos de conteúdo ainda se faz presente.

## 7. Discussão do estudo piloto

Para avaliar o rendimento da metodologia empregada, que visa obter uma compreensão dos estudantes sobre a natureza da aprendizagem de estratégias tradutórias, o exame das traduções produzidas é uma peça-chave.

Para a discussão da análise das traduções dos respondentes, fiz uso de conceitos sobre equivalência tradutória e textualização (Costa, 2005) e sobre os princípios de livre escolha e idiomático (Sinclair, 1991), que parecem estar em funcionamento em toda a produção textual.

Segundo Costa, todos os textos parecem ser dependentes de outros textos, mas um texto traduzido depende de um outro texto específico, de um modo bastante peculiar. O texto que o tradutor escreve será baseado numa mensagem que já existe em forma de texto em outra língua. Portanto, o texto original limita o novo texto de inúmeras maneiras, sendo a mais visível delas o fato de que o texto do tradutor deve ter um alto grau de semelhança com o seu correspondente original, para que seja reconhecido como tradução. Nos estudos de tradução, essa semelhança é denominada *equivalência.*

Esta ocorre quando um texto na língua fonte (TLF) e um texto na língua alvo (que aqui chamarei de traduzido, TT) são relacionáveis à maior parte das mesmas características relevantes. Sob esse ponto de vista, o TT é considerado uma entidade autônoma, mas que ao mesmo tempo, está intimamente relacionada com a sua fonte. O plano dependente do texto traduzido tem a ver com seu aspecto enquanto tradução e investigá-lo significa examinar o relacionamento de um dado par de línguas em nível sistêmico, de suas idiossincrasias gramaticais e lexicais, assim como de suas incongruências e incompatibilidades. Por outro lado, o plano autônomo do texto traduzido tem a ver com o seu aspecto enquanto texto, ou seja, com o modo como estão agrupadas as unidades (palavras, orações, parágrafos, etc.).

Conforme observa Costa:

> Quando lidamos com equivalência é necessário distinguir dois momentos ou facetas: uma delas é a equivalência dos itens no nível da oração, ou equivalência propriamente dita, e a outra é a equivalência supra-oracional, ou a equivalência (macro) textual. As escolhas abertas ao tradutor são restritas no primeiro caso e quase infinitas no segundo (Costa, 2005:27).

O problema da equivalência propriamente dita é, sem dúvida, central a qualquer tradução. Contudo, um aspecto não menos crucial no processo tradutório é a construção de um novo texto, seu processo e seu produto, denominado pelo autor de *textualização.* A textualização funciona quando o escritor vai do (macro) ideacional para o (macro) textual, ou seja, ele começa com algumas idéias e usa o seu acervo de palavras, regras gramaticais, padrões retóricos e experiências passadas como leitor e escritor e produz um texto que representa, apenas parcialmente, o que desejaria dizer. Ele parte do ideacional para chegar ao textual; ele começa com um conjunto de significados para produzir um texto que, por sua vez, constitui um novo conjunto de significados.

Costa, tomando como base as ideias de Hoey e Coulthard (apud Costa, 2005), salienta que, parece haver, em cada texto, blocos ideacionais, alguns deles mais importantes do que outros para o sentido global. Segundo Hoey:

> A primeira frase de um relatório costuma ter um status especial de frase resumo, por isso, seria possível argumentar que a capacidade da frase 1 de fazer sentido junto com as outras frases do texto é resultado desse *status* especial (29).

E conforme Costa:

> Identificar o bloco ideacional que resume o sentido geral do texto seria muito útil, já que, conforme Coulthard, o que precisamos inicialmente é um resumo do conteúdo ideacional (2005: 29).

É interessante notar, por exemplo, que, a primeira frase do texto *The multidisciplinary team approach to the treatment of bipolar disorder: an overview,* funciona como um resumo, e pela comparação e contraste das escolhas dos respondentes[17] (especialista, tradutora, R1 e R2, grupo A/CM, e R3 e R4, grupo B/SM), podemos ter uma ideia das principais decisões textuais de cada um deles. A partir daqui, explico a notação utilizada nos exemplos retirados dos textos originais e traduções dos respondentes. Em cada sequência, o segmento em itálico corresponde ao original em lingua inglesa. Os trechos em fonte normal, em azul e marrom são, respectivamente, os da especialista e da tradutora. Os trechos em cor laranja são do grupo A/CM e os dois últimos, em cor verde, os dos respondentes do grupo B/SM.

Vejamos a seguir:

*Bipolar disorder is a chronic and recurrent disorder, and many factors have been associated with its course and prognosis.*

O transtorno bipolar é um transtorno crônico e recorrente e muitos fatores têm sido associados com seu curso e prognóstico.

O transtorno bipolar é uma doença crônica e recorrente, e muitos fatores têm sido associados com o seu desenvolvimento e prognóstico.

Transtorno bipolar é um distúrbio crônico e recorrente e muitos fatores têm sido associados com seu curso e prognóstico (R1, A/CM).

---

[17] Textualizações, na íntegra, no final deste artigo.

O transtorno bipolar é uma doença crônica e recorrente e muitos fatores têm sido associados com o seu curso e prognóstico (R2, A/CM).

Transtorno bipolar é uma crônica e recorrente desordem, e muitos fatores têm sido associados com sua conduta e prognóstico (R3, B/SM).

Transtorno Bipolar é uma crônica e recorrente desordem, e muitos fatores têm sidos associados com a conduta e diagnóstico (R4, B/SM).

Nas traduções feitas, houve uma equivalência tradutória, em que informações relevantes do texto original foram preservadas. No entanto, como se observa acima, os respondentes que não construíram os mapas (em verde) apresentaram textualizações problemáticas quanto à gramática e ao léxico, ao traduzirem *chronic and recurrent disorder* por *crônica e recorrente desordem* e *course and prognosis* por *conduta e diagnóstico,* distanciando-se das traduções feitas pela especialista e pela tradutora, utilizadas como referência nesta pesquisa.

Como mencionado anteriormente, o mapa conceitual pode ser entendido como uma representação visual utilizada para partilhar significados, pois explicita como o autor entende as relações entre os conceitos enunciados. Parece-me que os respondentes que construíram os mapas (apresentados nos anexos 1 e 2), como recurso tradutório, conseguiram reter o ideacional do texto e, consequentemente, tornaram a frase resumo acima, mais compreensível.

Diferentemente do escritor do texto original, o tradutor é aquele tipo especial de escritor que cria o texto não a partir do seu próprio ideacional, mas a partir de outro texto. Ele reescreve um texto original, e é limitado não apenas pela gramática, pelos padrões lexicais de sua língua e pela habilidade como textualizador, mas também pelas restrições impostas pelo texto, pelo seu tom e conteúdo (Costa, 2005). Segundo o autor, as relações mais importantes se dão entre o ideacional, intertextual e textual que permeiam todos os tipos de texto em qualquer língua. Nesta discussão, levarei em consideração os aspectos ideacionais e textuais.

Textualizar implica diferentes dificuldades em diferentes aspectos. No ideacional, conforme mostrado por Coulthard (apud Costa, 2005), uma das principais dificuldades dizem respeito ao sequenciamento e à sinalização (gramática). Como o tradutor necessariamente trabalha numa sequência, ele naturalmente traduz

oração por oração, grupo por grupo e até mesmo palavra por palavra. Segundo Costa (2005), parece que a maioria dos tradutores trabalha com as menores unidades do discurso, preocupando-se mais com a frase ou a oração do que com a assimilação do ideacional do texto inteiro. Ao se utilizar dessa estratégia, o texto traduzido apresentará limitações de ordem macrotextual, como coesão e relações de correspondência. Importante salientar que isso pode ocorrer nas re-textualizações de diferentes gêneros textuais, como por exemplo, de uma obra literária ou de um artigo científico.

Tomemos, como exemplo, a segunda frase do resumo traduzido para o estudo piloto:

*Dysfunction in social, professional or family life has been correlated with poor outcomes and increased risk of relapse and recurrence, especially when the patient does not adhere to the treatment regimen.*

Disfunção na vida social profissional ou familiar tem sido correlacionada com pobres resultados e aumento do risco de recaída e recorrência, especialmente quando o paciente não adere ao tratamento.

Disfunções na vida social, profissional ou familiar foram correlacionadas com resultados pobres e aumento do risco de reincidência e recorrência, especialmente quando o paciente não adere ao programa de tratamento.

Em disfunção social, profissional ou vida familiar têm sido correlacionados a resultados pobres e aumentam o risco de recaída e reincidência, principalmente quando o paciente não adere ao regime de tratamento (R1, A/CM).

Disfunção na vida social, profissional ou vida familiar tem sido correlacionadas com resultados pobres e aumento do risco de recaída e recorrência principalmente quando o paciente não adere ao tratamento (R2, A/CM).

Disfunção na vida social, profissional ou familiar tem sido correlatado como pobres resultados e aumentado o risco de recaída e retorno (da doença), especialmente quando o paciente não segue o regime do tratamento (R3, B/SM).

Disfunção social, profissional ou na vida familiar têm sido correlacionados com insatisfatório resultado e o aumento do risco de recaída e recorrência, especialmente quando o paciente não conforma-se com o regime do tratamento (R4, B/SM).

Nessa frase, quando os fatores associados (citados na primeira) são explicitados, percebe-se que aqueles que construíram o mapa têm uma ideia mais precisa do conteúdo ideacional do texto, no que se refere às relações de correspondência, inclusive fazendo uso de colocações verbais como, por exemplo, aderir ao tratamento. Suas textualizações estão mais próximas às da especialista e da tradutora. O termo adhere, nesse contexto específico, tem adesão como seu equivalente em língua portuguesa, que significa não apenas fazer uso da medicação prescrita, mas se responsabilizar junto ao terapeuta pelos resultados do tratamento. A pergunta que fica é: Por que aqueles que não fizeram o mapa, não usaram o termo cognato (aderir), como se não tivessem entendido o conteúdo ideacional do texto?

Outro traço onipresente em qualquer textualização é a sinalização que diz respeito à gramática e que varia segundo os hábitos e as habilidades do textualizador, conforme apontado por Coulthard (apud Costa, 2005):

> ...a reunião de orações num texto em si cria relações entre as orações; mas uma decisão secundária do escritor é sinalizar ou não essas relações por realização lexical. Se ele escolhe não fazê-lo, permitirá a possibilidade de ambigüidade ou malentendido. (Ibid: 39).

Por exemplo, a R3 traduziu *recorrência* como um sinônimo de *recaída* quando, na verdade, *recorrência* significa que *a doença retorna muitas vezes*. Ao traduzir *recurrent* por *retorno da doença* e ao associar os conceitos de *recaída e retorno da doença,* a respondente dá uma ideia de que uma coisa leva à outra, mudando o significado da frase e, possivelmente, causando certa ambiguidade ao texto.

Importante lembrar que esta análise está sendo feita, levando em consideração a textualização de alunos de Letras em formação, que possuem pouca ou nenhuma experiência em tradução, como é o caso

do grupo B, que não construiu os mapas. Ao mesmo tempo, fica a pergunta: por que a palavra *recurrent* não foi traduzida por *recorrente*, a equivalência mais próxima na língua portuguesa? Será que é uma consequência da não assimilação do ideacional do texto inteiro, levando a uma preocupação apenas da frase? Será que os mapas conceituais que fornecem um entendimento do todo, partindo de proposições mais amplas para as mais específicas e organizando o conhecimento de forma estruturada, fez com que o grupo B, em alguns momentos, tivesse mais dificuldade em fazer uso de conceitos mais precisos?

Falemos, agora, sobre o que Sinclair (1991) chama de *princípio da livre escolha* e *princípio idiomático.* O autor defende a ideia de que, para explicar o modo como o significado surge a partir do texto linguístico, teremos que desenvolver dois princípios de interpretação diferentes.

O primeiro trata o texto linguístico como resultado de um número grande de escolhas complexas. Em cada ponto em que uma unidade está completa (uma palavra, um sintagma ou uma oração), abre-se uma ampla variedade de escolhas, e a única restrição é a gramaticalidade. Este princípio é formado por dois tipos de escolha: a lexical e a gramatical.

O segundo princípio diz que o usuário da língua dispõe de um grande número de sintagmas semi-pré-construídos que constituem opções unitárias, mesmo que possam parecer analisáveis em partes. Este é constituído de três conjuntos diferentes de escolhas mais amplas: a) expressões idiomáticas; b) colocações; c) sequências lexicogramaticais fixas.

Embora os dois princípios digam respeito às mesmas categorias linguísticas (léxico e gramática) a diferença essencial está no fato de que no princípio da livre escolha, a escolha é múltipla e, no principio idiomático a escolha é unitária.

Numa tradução literal, o princípio idiomático tende a desempenhar um papel menor ou a estar ausente, com expressões idiomáticas e colocações naturais, ocorrendo apenas quando houver correspondência biunívoca entre as palavras das línguas envolvidas.

Tomemos como exemplos, mais alguns trechos do resumo traduzido pelos respondentes, com base nos princípios da livre escolha

(lexical e gramatical) e do princípio idiomático (colocações e sequências fixas):

a) Princípio da livre escolha (lexical):

*The multidisciplinary team approach to the treatment of bipolar disorder: an overview*

A especialista não traduziu o título.

A abordagem da equipe multidisciplinar no tratamento do transtorno bipolar: uma visão geral.

A equipe de aproximação multidisciplinar para o tratamento do transtorno bipolar: **um panorama** (grifo meu) (R1, A/CM)

A abordagem multidisciplinar para o tratamento do transtorno bipolar: uma visão geral (R2, A/CM)

A Abordagem Multidisciplinar para o Tratamento do Transtorno Bipolar: uma visão geral. (R3, B/SM)

A abordagem multidisciplinar de uma equipe no tratamento do transtorno bipolar: uma visão geral. (R4, B/SM)

A escolha lexical da tradutora foi reproduzida, nas textualizações, por todos os respondentes, exceto pelo R1 (Grupo A/CM), que preferiu utilizar a palavra *panorama* (em negrito).

b) Princípio da livre escolha (gramatical):

*and increased risk of relapse and recurrence*

e aumento do risco de recaída e recorrência(especialista)

e aumento do risco de reincidência e recorrência(tradutora)

e aumentam o risco de recaída e reincidência (R1, A/CM)

e aumento do risco de recaída e recorrência (R2. A/CM)

e aumentado o risco de recaída e retorno (da doença) (R3, B/SM)

e o aumento do risco de recaída e recorrência (R4, B/SM)

Nessa passagem, a especialista, a tradutora, a R2 e a R4 optaram pelo uso do substantivo *aumento* na textualização, embora *increased,* nesta frase, seja um particípio, pois está relacionado com o tempo verbal presente perfeito na frase anterior: *Dysfunction in social, professional or family life has been correlated with poor outcomes and increased risk of relapse and recurrence.*

A R1 optou pelo presente simples *aumentam,* o que não é problemático para o entendimento do texto, pois o presente perfeito, neste caso, pode se referir à uma ação que ainda acontece.

A R3, que optou por *aumentado,* preservou o paralelismo gramatical sintático textualizado anteriormente na frase *tem sido correlatado como pobres resultados e aumentado o risco de recaída e retorno,* causando, possivelmente, uma certa estranheza ao leitor de língua portuguesa. Talvez isso tenha ocorrido devido à tentativa de utilizar o equivalente natural mais próximo de *correlated* e *increased.*

c) Princípio idiomático (colocações):

*especially when the patient does not adhere to the treatment regimen.*

especialmente quando o paciente não adere ao tratamento.

especialmente quando o paciente não adere ao programa de tratamento.

principalmente quando o paciente não adere ao regime de tratamento (R1, A/CM).

principalmente quando o paciente não adere ao tratamento (R2, A/CM).

especialmente quando o paciente não segue o regime do tratamento (R3, B/SM).

especialmente quando o paciente não conforma-se com o regime do tratamento (R4, B/SM).

Nesse trecho, a colocação *aderir ao tratamento,* como utilizada na linguagem médica (sentido explicado anteriormente) foi a escolha tradutória dos quatro primeiros respondentes: a especialista, a tradutora, a R 1 e a R 2 (grupo A/CM). A R3 e a R4 (grupo S/M) optaram por *seguir o regime de tratamento* e *conformar-se com o tratamento*, respectivamente. A opção da R3 não revela uma escolha problemática, ou seja, não interfere no entendimento do texto. A da R4, no entanto, ao utilizar o verbo *conformar-se,* modifica o sentido da frase, comprometendo o entendimento dessa parte do texto.

d) Princípio idiomático (sequências lexicogramaticais fixas):

*early identification of prodromal symptoms*

identificação precoce dos sintomas prodrômicos

identificação precoce dos sintomas prodrômicos

identificação precoce dos sintomas das etapas da doença (R1, A/CM)

identificação precoce dos sintomas prodrômicos (R2, A/CM)

precoce identificação de prodrômicos sintomas (R3, B/SM)
identificação precoce dos sintomas prodrômicos (R4, B/SM)

A sequência fixa no TT, correspondente à *identification of prodomical symptoms,* foi reproduzida por 4 dos 6 respondentes. A R1 traduziu por a *identificação precoce dos sintomas das etapas da doença,* mudando o significado da frase e comprometendo a informação veiculada pela mesma. A R3 optou por *precoce identificação de prodrômicos sintomas*, alterando a sequência lexicogramatical da expressão convencional *identificação de sintomas*, na língua portuguesa.

A análise feita acima, em fase inicial, intentou mostrar as diferentes condições de formulação de textualizações dos estudantes envolvidos, tendo em vista a familiarização de estratégias tradutórias para um grupo de sujeitos que não possuem um conhecimento experto em Tradução.

## 8. Síntese das observações e considerações finais

A partir da situação problema apresentada neste texto, que foi *Considerando a lacuna de formação existente relativa ao tema /natureza da Tradução em meio à formação do docente de LE, e que contribuir para diminuí-la é algo que pode ser positivo, como se pode introduzir esse tipo de conhecimento para o acadêmico de Licenciatura em Letras?* propus um modelo didático para a familiarização e a aprendizagem de estratégias tradutórias para alunos do curso superior de Licenciatura em Letras, que participaram como respondentes da minha pesquisa.

A metodologia utilizada no experimento mostrou resultados positivos no que se referem aos textos traduzidos por esses respondentes, no par inglês-português, considerando sua pouca ou nenhuma experiência com tradução.

Dessa forma, posso afirmar que é possível, sim, familiarizar esses estudantes com o tema da Tradução, através de uma metodologia centrada no aprendiz e baseada em tarefas, com níveis graduais de dificuldade, e adotando-se como estratégia cognitiva, a construção de mapas conceituais. Como vimos anteriormente, essa estratégia permite ao indivíduo estabelecer uma forma estável de

codificação de informações, que possibilita sua recuperação consciente por meio de redes associativas e a aquisição de habilidades determinantes para a compreensão da área temática a ser traduzida.

O propósito de incentivar o docente em língua inglesa em formação a desenvolver estratégias de tradução está imbuído do espírito de conscientizá-lo da possibilidade de tratamento desse tema paralelamente ao do ensino, como uma segunda possibilidade de formação profissional em Letras.

No que diz respeito aos dados iniciais obtidos, comparando-se os dois grupos (A/CM e B/SM), observa-se que o grupo que construiu e se utilizou dos mapas para as textualizações apresentou, de modo geral, produções mais próximas às da tradutora e da especialista, utilizadas como referência nessa pesquisa. Esses dados parecem indicar que o primeiro grupo fez uso de conceitos mais precisos, identificou o conteúdo ideacional do texto, apresentou menos problemas de sinalização e de colocação, em suas textualizações. Os resultados também indicam que a hipótese merece esse estatuto em meio ao trabalho da tese em andamento.

Por fim, acredito que uma análise detalhada sobre *como o aluno chega à boa solução na língua de chegada* e o *porquê de suas escolhas tradutórias* - por meio da descrição da travessia entre a compreensão da área temática, a aprendizagem das diferentes estratégias e o texto traduzido - se faz necessária futuramente, para que se possa desenvolver, com mais propriedade, uma pedagogia de tradução voltada para o professor de língua inglesa em formação.

## Referências

AUSUBEL, D. P. (1968). *Educational Psychology*: a cognitive view. New York: Holt, Rinehart and Winston, Inc.

AUSUBEL, D. P. (2000). "The Acquisition and Retention of Knowledge: A Cognitive View": Kluwer Academic Publishers.

CABRÉ, M. T. (1999c). *Terminología:* Representación y Comunicación. Elementos para uma Teoria de Base Comunicativa y otros Artículos. Sèrie Monografies, 3. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Linguística Aplicada.

COSTA, W. (2005). O Texto Traduzido como Re-textualização. *Cadernos de Tradução.* Vol. 2, no. 16. Santa Catarina: UFSC.

DARIN, L. C. de M. (2006). Questões Polêmicas nos Estudos da Tradução: foco no ensino-aprendizagem. *Estudos Lingüísticos,* XXXV. São Paulo: PUC-SP.

HURTADO-ALBIR, A. (2001). Caracterización de la Tradutología. *Traducción y Tradutología.* Madrid: Cátedra.

______. (2005). A Aquisição da Competencia Tradutória. In: *Competência em Tradução.* Belo Horizonte: UFMG.

MACHADO-VIEIRA, R.; SANTIN, A. e SOARES, J. C. (2004). The Multidisciplinary Team Approach to the Treatment of Bipolar Disorder: an overview. *Rev. Bras. Psiquiatr.* [online]. vol. 26, suppl.3, pp. 51-53.

NOVAK, J. D. (2010). *Learning, Creating and Using Knowledge.* Second Edition. New York: Routledge.

RODILLA, B. M.G. (1998). *La Ciencia empieza en la palabra.* Barcelona: Ediciones Península.

SINCLAIR, J. (1991). *Corpus Concordance Collocation.* Oxford: Oxford University Press.

TAVARES, R. *Construindo Mapas Conceituais*. Revista Ciências & Cognição (online), Vol 12: 72-85. Acessado em janeiro/2010, no sítio http://www.cienciasecognicao.org, 2007.

# ANEXOS

THB
sigla de
Transtorno de Humor Bipolar
transtorno caracterizado por
alternância e instabilidade de humor, em vários graus de apresentação, intensidade ou duração, com prejuízo das atividades familiares, sociais e/ou profissionais
com alternância de ciclos
de MANIA (euforia)
cujos sintomas são
+ 1 semana de humor elevado, distração, impaciência, grandiosidade, gastos imoderados, insônia, tagarelice, arriscar-se
e que são
tradicionalmente negligenciados como patológicos
de DEPRESSÃO
caracterizada por
+ 2 semanas humor irritado ou profundamente triste, sensação de vazio, desesperança, desânimo, medo, insegurança
MISTOS
que são
uma mistura de sintomasde mania e depressão
que exige
diagnóstico precoce e avaliação clínica minuciosa
adesão do paciente ao tratamento
com suporte
da rede familiar e social
que deve prover
atmosfera amistosa estímulo limites apoio
tratamento imediato
com
psicólogos, enfermeiros, nutricionistas, assitente social, grupo de apoio
além
do uso racional de medicamentos, por exemplo: estabilizadores de humor (carbonato de lítio)
cujos primeiros passos são
identificação das características da doença
que significa
separar o que é normal do ser humano do que é sintoma
avaliar impacto na vida do paciente e tempo de duração dos sintomas
determinar o significado atribuído à doença pela família e paciente
análise do contexto sócio-familiar do paciente (abordagem multicausal)
cuja causa pode ser
transmissão genética (substrato biológico)
cujos desencadeadores podem ser
estressores vitais psicológicos sociais
que classifica-se em
Tipo 1
que apresenta
1 ou + episódios de mania/ mistos
Tipo 2
isto é
1 ou + episódios depressivos + hipomaníacos (específicos da THB)
com diagnóstico estando entre
o THB tipo 1 e a Depressão Maior
Transtorno Ciclotímico
caracterizado por
numerosas ocorrência de hipomania e depressão em dois anos

# ANEXOS

Transtorno Bipolar
tem como causa
outros
transmissão genética
tem como fatores precipitantes estressores
vitais
sociais
psicológicos
é caracterizado pela alternância episódios
é um tipo de
transtorno de humor
tem
tratamento
é feito através de medicamentos como
estabilizadores de humor
anticonvulsionantes
para isso é necessário que
antipsicóticos
Depressão Maior
é tratado com
antidepressivos
sintomas como
isolamento social
desespero
medo, insegurança
pessimismo, baixa autoestima
outros
de euforia/Mania
hipomaniaco
para ser
mistos
por períodos de normalidade
sintomas como
humor eufórico
irritabilidade
impaciência
exaltação
outros
mistura de episódios
têm mais dificuldade na identificação
pode ter sucesso com
detecção
é
requisito
pode apresentar sintomas de
familiares e amigos estejam atentos
e devem encaminhar o portador a um
tratamento apropriado
evita-se
monoterapia
para
Bordeline
alta raiva
alta impulsividade
sintomas
menos depressão
falta de objetivos
baixa empatia
episódio maníaco de menor intensidade
em que o portador possui
relativo controle
divide-se
Tipo I
tem
ciclos bem marcados e espaçados
Tipo II
tem
maior instabilidade
oscilações rápidas e breves de humor
Tipo II e 1/2
diferenciam-se
TDAH
em
crianças
adultos
sobreposição de sintomas
tratamento
1º estabilização do humor
2º trata-se TDAH
ocorre
disfunção executivas
ocorre
labilidade afetiva
Nanashara Behle

# ANEXOS

## The multidisciplinary team approach to the treatment of bipolar disorder: an overview

Bipolar disorder is a chronic and recurrent disorder, and many factors have been associated with its course and prognosis. Dysfunction in social, professional or family life has been correlated with poor outcomes and increased risk of relapse and recurrence, especially when the patient does not adhere to the treatment regimen. Within the last decade, new treatments, intended to promote better adherence and minimize the risk of morbidity or hospitalization, have been tested. The multidisciplinary team approach attempts to educate patients and their families about such factors. Herein, we evaluate the therapeutic efficacy of this approach in applying the various psychosocial interventions employed in the treatment of bipolar disorder. The objective of this approach is early identification of prodromal symptoms in order to prevent hospitalization and behavioral dysfunction.

**Respondente 1 (a especialista):**

O transtorno bipolar é um transtorno crônico e recorrente e muitos fatores têm sido associados com seu curso e prognóstico. Disfunção na vida social profissional ou familiar tem sido correlacionada com pobres resultados e aumento do risco de recaída e recorrência, especialmente quando o paciente não adere ao tratamento. Na última década novos tratamentos têm sido testados na intenção de promover melhor aderência do paciente ao tratamento e minimizar o risco de morbidade ou hospitalização. A abordagem da equipe multidisciplinar tenta educar os pacientes e seus familiares sobre tais fatores. Aqui nós avaliamos a eficácia terapêutica da nossa abordagem, aplicando varias intervenções psicossociais no tratamento do transtorno de humor bipolar. O objetivo dessa abordagem é validar a identificação precoce dos sintomas prodrômicos na prevenção da hospitalização e do comportamento disfuncional

**Respondente 2 (a tradutora):**

**A abordagem da equipe multidisciplinar no tratamento do transtorno bipolar: uma visão geral**

O transtorno bipolar é uma doença crônica e recorrente, e muitos fatores têm sido associados com o seu desenvolvimento e prognóstico. Disfunções na vida social, profissional ou familiar foram correlacionadas com resultados pobres e aumento do risco de reincidência e recorrência, especialmente quando o paciente não adere ao programa de tratamento. Na última década, novos tratamentos foram testados a fim de promover uma melhor adesão do paciente ao tratamento e minimizar o risco de morbidade ou hospitalização. A abordagem da equipe multidisciplinar intenta educar os pacientes e suas famílias sobre tais fatores. Neste artigo, nós avaliamos a eficácia terapêutica desta abordagem ao aplicarmos as várias intervenções psicossociais empregadas no tratamento do transtorno bipolar. O objetivo da abordagem que aqui nos referimos é a identificação precoce dos sintomas prodrômicos para que se possa prevenir a hospitalização e a disfunção comportamental.

**Respondente 1 (Grupo A/CM): A equipe de aproximação multidisciplinar para o tratamento do transtorno bipolar: um panorama**

Transtorno bipolar é um distúrbio crônico e recorrente e muitos fatores têm sido associados com seu curso e prognóstico. Em disfunção social, profissional ou vida familiar têm sido correlacionados a resultados pobres e aumentam o risco de recaída e reincidência, principalmente quando o paciente não adere ao regime de tratamento. Dentro da última década, novos tratamentos, com a intenção de promover uma melhor aderência e minimizar o risco de morbidez ou hospitalização, têm sido testados. A equipe de aproximação disciplinar tenta educar os pacientes e seus familiares sobre estes fatores. Incluindo, nós avaliamos a eficácia terapêutica desta aproximação na aplicação de várias intervenções psicosociais aplicadas ao tratamento do transtorno bipolar. O objetivo dessa aproximação é a identificação precoce dos sintomas das etapas da doença a fim de prevenir a hospitalização e a disfunção comportamental.

**Respondente 2 (A/CM): A abordagem multidisciplinar para o tratamento do transtorno bipolar: uma visão geral.**

O transtorno bipolar é uma doença crônica e recorrente e muitos fatores têm sido associados com o seu curso e prognóstico. Disfunção na vida social, profissional ou vida familiar tem sido correlacionadas com resultados pobres e aumento do risco de recaída e recorrência principalmente quando o paciente não adere ao tratamento. Na última década, foram testados novos tratamentos destinados a promover uma melhor aderência e minimizar o risco de morbidade ou de internação. A abordagem multidisciplinar para o transtorno bipolar busca educar os pacientes e suas famílias sobre tais fatores. Nesse trabalho, nós avaliamos a eficácia terapêutica desta abordagem através da aplicação das várias intervenções psicossociais empregadas no tratamento do transtorno bipolar. O objetivo desta abordagem é a, a fim de evitar a hospitalização e disfunção comportamental.

**Respondente 3 (Grupo B/SM): A Abordagem Multidisciplinar para o Tratamento do Transtorno Bipolar: uma visão geral**

Transtorno bipolar é uma crônica e recorrente desordem, e muitos fatores têm sido associados com sua conduta e prognóstico. Disfunção na vida social, profissional ou familiar tem sido corelatado como pobres resultados e aumentado o risco de recaída e retorno (da doença), especialmente quando o paciente não segue o regime do tratamento. Na última década, novos tratamentos, destinados a promover melhores aderências e minimizar o risco de morbidade ou hospitalização, têm sido testados. As tentativas de abordagem multidisciplinar de educar pacientes e suas famílias sobre tais fatores. Nisto, nós avaliamos a eficácia terapêutica dessa abordagem na aplicação de várias intervenções psicossociais empregadas no tratamento do transtorno bipolar. O objetivo dessa abordagem é a precoce identificação de prodrômicos sintomas a fim de prevenir a hospitalização e a disfunção comportamental.

**Respondente 4 (Grupo B/SM): A abordagem multidisciplinar de uma equipe no tratamento do transtorno bipolar: uma visão geral**

O transtorno bipolar é uma doença crônica e recorrente, e muitos fatores têm sido associados com o seu curso e prognóstico. Disfunção na vida social, profissional ou familiar tem sido correlacionada com resultados ruins e o aumento do risco de recaída e recorrência, principalmente quando o paciente não adere ao tratamento. Na última década, novos tratamentos tentaram promover uma melhor aderência e minimizar o risco de morbidade ou de internação, foram testados. A abordagem multidisciplinar tenta educar os pacientes e suas famílias sobre tais fatores. Nisto, nós avaliamos a eficácia terapêutica desta abordagem na aplicação das várias intervenções psicossociais empregadas no tratamento do transtorno bipolar. O objetivo desta abordagem é a identificação precoce dos sintomas prodrômicos, a fim de evitar a hospitalização e disfunção comportamental.

# TERMINOLOGIA TEXTUAL E LINGUÍSTICA DE *CORPUS*: ESTUDO EM PARCERIA

*Leonardo Zilio*[1]

***Resumo**: Este texto mostra algumas possibilidades de cooperação entre a Terminologia Textual e a Linguística de Corpus. Nosso primeiro passo é descrever um pouco as áreas abordadas, principalmente por serem áreas de estudo relativamente novas. No caso da Terminologia Textual, há ainda toda uma problematização em relação à sua concepção e existência como área de estudo, a qual não pode passar em branco. Para mostrar como essas áreas podem auxiliar-se mutuamente, apresentamos, posteriormente, dois trabalhos já realizados. O primeiro trabalho diz respeito à caracterização de gêneros e tipos textuais. Já o segundo trabalho entra no ramo da Fraseologia Especializada e se ocupa de colocações especializadas. Ambos os trabalhos foram realizados com textos da área da Cardiologia. Como se pode ver, apesar de mencionarmos inicialmente apenas duas áreas de estudo, a Terminologia Textual e a Linguística de Corpus, a interdisciplinaridade é muito maior, abrangendo diversas áreas.*

## 1 Terminologia Textual e Linguística do Texto Especializado

Quando mencionamos a expressão *Terminologia Textual*, não podemos deixar de esclarecer o que estamos entendendo sob essa denominação. Isso porque, não bastasse a Terminologia já apresentar algumas controvérsias[2], o título **Terminologia Textual** já foi utilizado para descrever outros tipos de estudo que não são abordados aqui.

---

[1] Mestre em Estudos da Linguagem pelo PPG-Letras da UFRGS na linha de pesquisa Lexicografia e Lexicografia: Relações Textuais.

[2] O termo **Terminologia** permite fazer referência tanto à área de estudos como ao conjunto de termos de uma área. Além disso, a área de estudos pode ser entendida como eminentemente normativa ou descritiva, dependendo da abordagem que se toma como ponto de partida.

Condamines (2005), por exemplo, propõe que, na Terminologia Textual, os *corpora* sejam utilizados para auxiliar o terminólogo na busca por **termos**. Krieger (2008, p. 6), por outro lado, adota um ponto de partida baseado na semiótica greimasiana e deixa claro que, para ela, o termo "Terminologia Textual está relacionado à integração de componentes de textualidade e da discursividade no aparato teórico-metodológico da Terminologia, **cujo objeto primeiro é o termo técnico-científico**" (grifo nosso). Veremos mais adiante que o que estamos entendendo por Terminologia Textual é mais abrangente do que essas propostas.

Os estudos terminológicos e as teorias de Terminologia, começando por Wüster, com a Teoria Geral da Terminologia (1974), passando pela Teoria Comunicativa da Terminologia de Cabré (2001) e pela Teoria Sociocognitiva de Temmermann (2000), apresentam-se acentuadamente, ou mesmo totalmente, vinculados ao estudo dos termos. Cada autor apresenta uma perspectiva diferente, embora a base de pesquisa sempre esteja associada àquelas unidades que denotam conceitos especializados.

Nos anos 80, porém, começaram a se desenvolver estudos, quer se denominassem terminológicos ou não, que já compreendiam a importância do texto especializado. Leitchik (2004), em uma coletânea de artigos russos sobre Terminologia do período de 1992 a 2002, apresentou uma proposta de trabalho que denominou de **Teoria Terminológica do Texto** (*Terminological Text Theory*), a qual já vinha desenvolvendo desde 1981 (cf. Leitchik, 2004, p. 373). A proposta de Leitchik ainda está vinculada ao conceito de termo e de sua centralidade para o estudo terminológico. Todavia, já é reconhecido pelo autor que o objeto de estudos da Teoria Terminológica do Texto compreende vários **textos, nos quais termos estão contidos**[3] (cf. Leitchik, 2004, p. 371).

---

[3] Essa oposição do estudo de termos em textos e de textos com termos é abordada por Finatto (2004a) em um artigo muito interessante que apresenta, entre outros, uma boa introdução para o que está sendo discutido nesta seção.

Os estudos de textos que contêm termos se desenvolveram principalmente no mundo germânico, onde o estudo da *Fachsprache*[4] (Finatto, 2004a, p. 347), principalmente apoiada nos estudos e propostas de Hartwig Kalverkämper (1983) e Lothar Hoffmann (1988a), aborda o texto especializado como elemento central do estudo das linguagens especializadas. Esses autores não se referem aos seus estudos pelo título Terminologia Textual, mas sim como Linguística de Linguagens Especializadas ou Linguística do Texto Especializado. Uma explicação para isso é que, na época, o paradigma da Terminologia era representado pela proposta de Eugen Wüster, bem sintetizada em seu artigo póstumo de 1974.

Contribui para essa explicação o fato de que o próprio Hoffmann (1988a) não dizia estar fazendo estudos terminológicos, mas sim estudos na área da *Fachsprachenforschung* [Pesquisa de linguagens especializadas], que ele diferenciava do *Terminologiearbeit* [Trabalho terminológico], entendido como a área que se encarregava do estudo dos termos a partir de uma perspectiva normativa.

O que nos fascina nos trabalhos de Hoffmann e Kalverkämper é a proposta de se estudarem elementos que não dependem diretamente do termo. Eles entendem que, do ponto de vista linguístico, não há como algo que se pretende como língua ou linguagem ser apenas composto pelo léxico e, apesar de não se oporem diretamente à teoria de Wüster, percebe-se que a sua posição não é favorável ao que está sendo feito:

> O trabalho terminológico é geralmente compreendido como uma parte da linguística de linguagens especializadas, mas na verdade ele se ocupa apenas de um núcleo das linguagens especializadas – se ocupa da terminologia. Esse trabalho se situa num lugar bastante específico, sobretudo porque, em

[4] Esse termo, em alemão, permite a tradução tanto por língua de especialidade quanto por linguagem de especialidade, ou, ainda, por linguagem especializada. Na nossa compreensão dos fatos, acreditamos que, dependendo do contexto, possa fazer referência ou à língua de especialidade ou à linguagem especializada. Para melhor compreender a distinção que fazemos entre língua de especialidade e linguagem especializada, aconselhamos a leitura de Zilio (2009), principalmente do capítulo 1.

> grande parte, nele não estão envolvidos lingüistas, mas apenas engenheiros e técnicos. Essa situação ainda deverá se prolongar enquanto não haja condições favoráveis para a linguística de linguagens especializadas nessas instituições. (Hoffmann, 1988b, p. 87)

Sua proposta, então, é que os termos sejam somente um dos aspectos estudados nas linguagens especializadas. Hoffmann se apega à Linguística de Linguagens Especializadas, à qual Kalverkämper (1983, p. 125) define como "teoria terminológica com uma série de problemáticas adicionais"*, ou seja, como algo que abrange mais do que o léxico, considerado frequentemente como o "*Hauptträger der Fachsprachlichkeit*" [principal portador da especialização linguística] (cf. Kalverkämper, 1983, p. 125).

> Mas, assim como as unidades lexicais estão alojadas em um todo de ordem maior, nominadamente, a oração, também a oração é uma parte completa de um todo de ordem maior e ainda mais complexo, para cuja constituição ela contribui, nominadamente, o texto* (Kalverkämper, 1983, p. 126).

Assim, o paradigma proposto tem como objeto central de estudo o texto especializado e não mais a unidade terminológica. Hoffmann (1998a, p. 416) aponta que "o específico das linguagens especializadas se expressa mais visivelmente em seu vocabulário"*, porém, fica claro que essa especificidade não está somente no vocabulário. Isso representa uma grande ampliação dos estudos relacionados às linguagens especializadas, uma ampliação bastante coerente e importante, já que, se observarmos atentamente, um estudo somente sobre termos não esgota os questionamentos restantes acerca de uma linguagem especializada.

---

* No original: "Terminologielehre mit einem zusätzlichen Fächer von Problemstellungen".

* No original: "Aber so wie die lexikalischen Einheiten eingebettet sind in ein übergeordnetes komplexeres Ganzes, nämlich in den Satz, so ist der Satz selbst ein ganzheitlicher Teil eines wiederum übergeordneten, noch komplexeren Ganzen, zu dessen Konstituierung er beiträgt, nämlich des Textes".

* No original: "die Spezifik der Fachsprachen äußert sich besonders deutlich in ihren Wortschätzen".

Apresenta-se então a ***Fachtextlinguistik***[5] (cf. Kalverkämper, 1983). Esse termo, em alemão, pode ser interpretado de duas maneiras: de um lado, temos a ***Fach-Textlinguistik***, que, buscando responder o que é uma especialidade, se ocuparia mais especificamente da pragmática, “pois a busca pela especialidade ou especialidades visa às ações humanas”[*] (Kalverkämper, 1983, p. 128); por outro lado, temos a ***Fachtext-Linguistik***, entendida como uma Linguística de Gêneros Textuais e que se ocupa dos textos em funcionamento.

É essa segunda interpretação que nos interessa aqui de forma mais específica. Seu objetivo é descrever o funcionamento do texto, seja seu funcionamento interno ou externo, de forma que traz para a análise também os participantes da comunicação. Apontar como funciona o gênero textual e como ele se distingue de outros gêneros textuais faz parte de um estudo do texto especializado. Para isso, utilizam-se métodos empíricos de análise, que buscam identificar os pontos que o fazem especializado e o distinguem. Tendo-se o texto como foco, deixa de fazer sentido que se continue estudando somente os termos, de forma que se passa a englobar “os modos de dizer peculiares de cada área de conhecimento” (Finatto, 2004a, p. 348).

Essa distinção do texto como objeto de estudo aparenta ser uma proposta bastante ambiciosa, porém, deve-se levar em conta que ele é o objeto da Linguística do Texto Especializado, a qual engloba várias outras áreas de estudo que têm como objeto as diferentes unidades presentes em um texto especializado. O todo “se constrói a partir de fonemas, passando por palavras, sintagmas, orações, sequências oracionais, seções/capítulos, texto, comunidades textuais, até gêneros literários (ou gêneros textuais), o que já não pode mais ser observado

---

[5] *Fachtextlinguistik*, em alemão, se não tiver nenhum sinal gráfico que aponte qual a divisão que se deve fazer na palavra, pode ser interpretada de duas formas. A primeira forma *Fach-Textlinguistik*, seria uma Linguística Textual Especializada; já a segunda forma possível, *Fachtext-Linguistik*, seria uma Linguística do Texto Especializado. A diferença entre as duas formas é mais aprofundada em Kalverkämper (1983). Neste estudo, elas são abordadas somente para apontar as diferenças mais básicas.

[*] No original: “denn die Frage nach dem Fach oder den Fächern zielt auf das menschliche Handeln”.

como uma completude"[*] (Kalverkämper, 1983, p. 153). Hoffmann (1988a, p. 22), abordando a Pesquisa de Linguagens Especializadas, apresenta-nos uma subdivisão em cinco níveis decrescentes: texto (textemas), frase (sintagmas), palavra (lexemas ou sememas), forma (morfemas), letras ou sons (grafemas ou fonemas); e aponta que cada uma dessas áreas pode ser estudada a partir de cinco métodos de estudo: estrutural, semântico, estatístico, comparativo e funcional.

Tendo passado por esse quadro teórico, cremos que está mais claro o porquê de chamarmos essa área de estudo de Terminologia Textual. Sabemos que ainda persistem problemas em relação a essa denominação, já que o termo **Terminologia**, aponta, filologicamente, para um estudo do termo. Por essa mesma perspectiva, fica também estranho entender que algo **terminol**ógico possa ser, ao mesmo tempo, **textual**. Essa discussão não se encerra aqui, mas também não é em um artigo que ela se esgota e nem é esse o objetivo aqui estabelecido. O que deve ficar claro desde já é que a proposição de uma Terminologia Textual ainda é algo bastante recente no Brasil[6] e merece maior discussão.

Tendo apresentado um pouco das bases do que entendemos por Terminologia Textual, passamos à nossa outra área de estudos, a Linguística de *Corpus*.

## 2. Linguística de *Corpus*

A Linguística de *Corpus* é uma área da Linguística que vem ganhando cada vez mais força no mundo acadêmico e que tem desenvolvido trabalhos muito importantes no que diz respeito à descrição de língua e linguagem.

---

[*] No original: "baut sich auf aus den Phonemen über die Wörter, Syntagmen, Sätze, Satz-Folgen, Abschnitte/Kapitel, Text, Texte-Gemeinschaft, bis in die Gattungen (oder Textsorten) hinein, was als Ganzheit schon kaum mehr überschaubar ist".

[6] Como referência nessa área de estudos, podemos mencionar os trabalhos desenvolvidos no âmbito do Projeto TEXTQUIM (em breve: TEXTECC; www.ufrgs.br/textquim). Aos poucos também se concretiza a criação de um círculo de estudos, sob a coordenação da Profª. Drª. Maria José Bocorny Finatto, para debater o assunto.

O grande desenvolvimento da Linguística de *Corpus* nos últimos anos se deve muito à (r)evolução informática. É claro que já existiam trabalhos realizados com *corpora* antes dos computadores. Porém, a partir do momento em que se pôde armazenar textos em um computador e produzir programas capazes de analisar esses textos de forma automática ou assistida por computador é que os trabalhos com *corpora* se tornaram mais frequentes, criando e dando maior visibilidade à área.

Devido à possibilidade de se aplicar, em outras áreas, métodos e ferramentas desenvolvidos dentro do âmbito da Linguística de *Corpus*, existem muitas dúvidas quanto ao estatuto dessa área de estudos: alguns autores acreditam ser ela somente uma metodologia, enquanto outros defendem que ela é, de fato, uma disciplina. Como nos aponta Berber Sardinha (2004, p. 36), se entendermos que uma metodologia é o instrumental aplicado em uma pesquisa, então podemos entender a Linguística de *Corpus* como tal, já que outras teorias podem recorrer das ferramentas desenvolvidas para o estudo de *corpora* para seus próprios fins. Porém, a Linguística de *Corpus* também conta com um aporte teórico além das ferramentas das quais faz uso. Seguindo essa ideia, Rajagopalan (2007, p. 33) escreveu, no início do livro *Um percurso para pesquisas com base em corpus*, "quem acha que a Linguística de *Corpus* não passa de uma ferramenta e nada ou pouco contribui para repensar os fundamentos da Linguística (…) está prestes a ser saudado com uma bela surpresa".

Uma terceira opção apontada por Berber Sardinha (2004, p. 37) é a compreensão da Linguística de *Corpus* como uma abordagem. Segundo Hoey (1997, apud Berber Sardinha, 2004, p. 37), "Linguística de *Corpus* não é um ramo da linguística, mas a rota para a linguística". Biber et alii (1998) também adotam essa terceira posição, denominando seu estudo de abordagem baseada em *corpus* e propondo quatro principais características para essa abordagem:

> - é empírica, analisando os padrões de uso existentes em textos naturais;
> - utiliza uma coleção grande e selecionada de textos naturais, conhecida como *corpus*, como base para a análise;

> -faz amplo uso de computadores para a análise, utilizando tanto técnicas automáticas como interativas;
> - depende de técnicas de análise tanto quantitativas quanto qualitativas[*] (Biber et alii, 1998, p. 4).

Nessa discussão, nos colocamos como adeptos da terceira opção, já que nosso aporte teórico, além de estar baseado na Linguística de *Corpus*, também está de mãos dadas com outras áreas da linguística, de forma que não podemos enquadrar nosso trabalho somente sob o teto de uma delas. Dessa forma, a Linguística de *Corpus* será a porta de entrada para os estudos linguísticos que intentamos, sendo seus pressupostos teóricos acolhidos.

Segundo Berber Sardinha (2004, p. 30), os dois principais pressupostos da Linguística de *Corpus* são:

- a linguagem é um sistema probabilístico de combinatórias; e
- a abordagem deve ser empírica.

Comecemos pela observação do segundo pressuposto. “Na linguística, empírico significa primazia aos dados provenientes da observação da linguagem, em geral reunidos sob a forma de um *corpus*” (Berber Sardinha, 2004, p. 30). Percebe-se, portanto, que a Linguística de *Corpus* está muito ligada à observação do que comumente se chama de **uso**.

Biber et alii (1998, p. 1) apresentam, logo no início de seu livro *Corpus Linguistics: Investigating Language Structure and Use*, uma diferenciação entre *language structure* [estrutura da linguagem] e *language use* [uso da linguagem]. Os autores mencionam claramente que seu foco é o uso: “estudamos a linguagem realmente utilizada em textos naturais”[*] (Biber et alii, 1998, p. 1).

Isso traz à tona uma questão importante: o que é **uso**? Stubbs (2001, p. 1) escreve que “a maioria das palavras utilizadas no dia a dia

---

[*] No original: “- it is empirical, analyzing the actual patterns of use in natural texts; - it utilizes a large and principled collection of natural texts, known as a ‘corpus’, as the basis for analysis; - it makes extensive use of computers for analysis, using both automatic and interactive techniques; - it depends on both quantitative and qualitative analytical techniques”.

[*] No original: “we study actual language used in naturally occurring texts”.

tem diferentes usos e diferentes significados"[*], dando a entender que *uso* e *significado* não são sinônimos. Porém, um dos lemas frequentemente utilizados pelo autor é justamente "meaning is use" [significado é uso] (cf. Stubbs, 2001, p. 13), mesmo admitindo que seja uma forma simplificada de se entender a sua proposta de trabalho (cf. Stubbs, 2001, p. 20). Stubbs (2001) trabalha com frequência, mas considera que os diferentes significados não são atrelados à frequência, mas sim aos diferentes contextos e às diferentes combinatórias – principalmente a essas últimas – em que uma palavra ou lema ocorre e esses fatores são identificados como uso. Portanto, uso é entendido não como ocorrência, mas sim como ocorrência em diferentes combinatórias.

Todavia, não se deve entender que a Linguística de *Corpus* está interessada somente na observação da linguagem. Stubbs chama atenção para essa generalização no seguinte trecho:

> Enquanto muitos lingüistas (chomskianos) têm se ocupado do que os falantes podem dizer, a Linguística de *Corpus* também está necessariamente ocupada com o que os falantes efetivamente dizem. Mas que fique marcado o *também*. É um equívoco ver somente freqüência de ocorrências reais. Freqüência se torna interessante quando é interpretada como algo típico, e a competência comunicativa dos falantes inclui conhecimento tácito de normas de comportamento[*] (Stubbs, 2001, p. 61).

Voltando aos dois pressupostos citados, vejamos o que se fala do primeiro. No que diz respeito à ideia de sistema probabilístico, tem-se que, "embora muitos traços lingüísticos sejam possíveis teoricamente, não ocorrem com a mesma freqüência" (Berber Sardinha, 2004, p. 30-31). Pelas recém-mencionadas palavras de Stubbs, porém, vemos que essa diferença de frequência não é aleatória, mas sistemática. Assim, se

---

[*] No original: "most everyday words have diferent uses and diferent meanings".

[*] No original: "Whereas much (Chomskyan) linguistics has been concerned with what speakers can say, corpus linguistics is also necessarily concerned with what speakers do say. But note the *also*. It is misleading to see only frequency of actual occurrence (…). Frequency becomes interesting when it is interpreted as typicality, and speakers' communicative competence includes tacit knowledge of behavioural norms".

pode dizer que a linguagem é padronizada. E essa padronização "se evidencia pela recorrência, isto é, uma colocação, coligação ou estrutura que se repete significativamente mostra sinais de ser, na verdade, um *padrão* lexical ou léxico-gramatical" (Berber Sardinha, 2004, p. 31). Esses padrões representam, segundo Biber et alii (1998, p. 5), "os caminhos sistemáticos em que os aspectos linguísticos são usados em associação com outros aspectos linguísticos e não-linguísticos"*. Por associações linguísticas, Biber et alii (1998, p. 6) entendem as associações lexicais e gramaticais; por não-linguístico, entende-se a observação da distribuição desses aspectos linguísticos em diferentes tipos de registro, dialetos ou períodos.

Foi baseado nessa padronização que John Sinclair (1991, p. 110-115) propôs o princípio idiomático da linguagem, pelo qual, "o usuário de uma língua tem ao seu dispor um grande número de sintagmas semi ou pré-construídos, que constituem em escolhas únicas, mesmo que pareçam ser analisáveis em segmentos". Esse princípio implica no fato de que a existência ou não de uma palavra em um determinado sintagma dependa de palavras selecionadas anteriormente. Assim, por exemplo, não é qualquer verbo que pode preencher a necessidade imposta por um sujeito, mas sim somente um determinado verbo ou grupo de verbos restringidos por ele.

Por investigar a linguagem através de uma massa de, muitas vezes, milhões ou mesmo bilhões de palavras, a Linguística de *Corpus* se apoia, hoje em dia, na investigação automática ou semiautomática de *corpora* armazenados em computadores. Existem muitas ferramentas já desenvolvidas para diversos fins.

O que há de mais interessante na Linguística de *Corpus* é que, por ser uma investigação baseada em *corpora*, pode-se deixar que os textos nos guiem. É a partir dos dados averiguados no *corpus* que um linguista de *corpus* pode conduzir sua pesquisa. O *corpus* não é somente um instrumento de confirmação de dados ou de suposições, mas sim uma fonte destes. E é isso o que mais nos atrai para nos voltarmos a esse tipo de abordagem.

---

* No original: "the systematic ways in which linguistic features are used in association with other linguistic and non-linguistic features".

## 3. Associando a Terminologia Textual à Linguística de *Corpus*

Trabalhando com duas áreas interdisciplinares, não é difícil perceber como elas podem cooperar. Nesta seção, apresentamos, muito rapidamente, alguns pontos de convergência ou de empréstimo entre as duas áreas que ficarão mais salientes ao observarmos resultados de estudos já realizados nessa parceria.

Em uma primeira abordagem, vemos que a Terminologia Textual entra com os métodos e aportes teóricos para o trabalho com textos especializados, enquanto a Linguística de *Corpus* entra com o aparato informático, com os pressupostos teóricos e com seu modo peculiar de observar a linguagem.

O trabalho com *corpora* também tem toque das duas áreas, já que a Linguística de *Corpus* aponta os melhores modos de organizar um conjunto de textos, enquanto a Terminologia Textual aponta os tipos de texto a serem recolhidos.

É também a partir da observação da linguagem especializada em *corpora* que se percebe mais francamente como a observação de termos é somente um pequeno passo na observação do texto especializado, mesmo que se entenda que esse é um passo importante na caminhada.

Assim, buscando elementos de uma área e da outra, se molda um tipo de estudo que já mostra seus frutos. Uma parte dos frutos que já foram colhidos é o que veremos nas seções seguintes deste capítulo.

## 4. Estudos em parceria

Nesta seção, apresentamos dois estudos que desenvolvemos, sendo o primeiro deles em colaboração com outros autores. Os trabalhos que apresentamos têm a característica comum de, como enfatizamos neste capítulo, mesclar duas abordagens em uma parceria produtiva.

O primeiro estudo dá conta de uma observação no âmbito dos gêneros textuais, procurando estabelecer semelhanças e diferenças entre tipos textuais diferentes e entre línguas diferentes. Já no segundo estudo, o assunto abordado e bastante diferente, ocupando-se de colocações especializadas em alemão e de seus equivalentes em português.

Ambos os trabalhos se desenvolveram principalmente a partir de um *corpus* formado por artigos científicos da área da Cardiologia escritos originalmente nas duas línguas supracitadas.

É importante ressaltarmos que, apesar de estarem sendo apresentados individualmente, ambos os estudos foram desenvolvidos em parceria entre os autores, durante o período de 2007 e 2009.

### *4.1. Gêneros Textuais*

O primeiro estudo a ser apresentado diz respeito a uma proposta de delimitação entre tipos textuais e foi desenvolvido paralelamente em dois estudos (Scheeren et alii, 2008; Zilio, 2009). Resultados foram posteriormente juntados e podem ser encontrados em Zilio et aliae (2009).

Esse estudo tinha por objetivo mostrar para tradutores iniciantes como textos de subáreas da Medicina podem apresentar características suficientemente diferentes para serem considerados tipos textuais distintos e como existem diferenças importantes entre duas línguas, visto que o estudo se deu entre o alemão e o português.

As orientações teóricas relativas aos gêneros textuais foram extraídas de Ciapuscio (1994; 2003) no que diz respeito à divisão entre classes e tipos textuais, advinda da tradição textual alemã. Podemos dizer, sinteticamente, que as classes textuais são consideradas como os gêneros textuais, são percebidas empiricamente, enquanto os tipos textuais são estabelecidos através de estudos baseados em uma teoria linguística. Essa teoria linguística foi encontrada em Hoffmann (1988a), que propõe que o estabelecimento entre os diferentes tipos textuais deve ser fundamentado em uma análise que contemple as unidades linguísticas presentes nos textos desde as menores, os grafemas, até as macroestruturas, passando pelos demais níveis microestruturais (morfológico, léxico e sintático) nesse caminho.

Para levar a cabo essa análise de diversos níveis textuais, recorreu-se à Linguística de *Corpus*, que, através de sua visão probabilística da linguagem, permite observar as preferências lexicais e sintáticas em um *corpus*.

O estudo comparou, primeiramente, textos de Cardiologia em alemão com textos de Radiologia em alemão e, posteriormente, textos de Cardiologia em alemão com textos de Cardiologia em português.

Tanto no estudo interlingual quanto no estudo entre as duas subáreas, os principais resultados surgiram na observação das microestruturas. As macroestruturas tendiam muito a seguir orientações dos periódicos estudados e não necessariamente vinculadas às áreas estudadas.

No contraste entre as subáreas, observou-se a presença de substantivos muito diferentes nas listas de palavras[7] (ver tabela 1) das duas áreas, tendo em comum somente *Patienten* [pacientes], *Ergebnisse* [resultados] e *Untersuchung* [análise].

Tabela 1. Os 10 substantivos mais frequentes em Cardiologia e Radiologia.

| **Cardiologia** | | | **Radiologia** | | |
|---|---|---|---|---|---|
| **Posição na lista** | Substantivo | Frequência relativa (%)[8] | Posição na lista | Substantivo | **Frequência relativa(%)** |
| **06** | Patienten | 1,37 | 11 | Patienten | **0,77** |
| **40** | Studie | 0,24 | 47 | Läsionen | **0,20** |
| **48** | Ergebnisse | 0,19 | 59 | Untersuchung | **0,17** |
| **56** | Jahre | 0,16 | 61 | Kategorie | **0,16** |
| **58** | Therapie | 0,16 | 63 | Ergebnisse | **0,16** |
| **61** | Gruppe | 0,16 | 64 | Fällen | **0,16** |
| **65** | Operation | 0,16 | 67 | Euro | **0,15** |
| **66** | Alter | 0,15 | 79 | Befunde | **0,13** |
| **67** | Studien | 0,15 | 81 | Untersuchungen | **0,12** |
| **70** | Untersuchung | 0,14 | 88 | Veränderungen | **0,11** |
| **Total** | | **2,88** | **Total** | | **2,13** |

[7] As listas de palavras foram geradas a partir do software *Wordsmith Tools*, utilizando-se a ferramenta *WordList*.

[8] Optamos por apresentar somente as frequências relativas, visto que as frequências absolutas podem ser obtidas a partir desse valor, aplicando-o em relação ao tamanho total do *corpus* (disponível na seção 5.1).

Após terem sido observadas essas listas, partiu-se para uma observação de associações lexicais, mostrando que, no nível sintático, as diferenças são ainda maiores. Para essa observação de associações, foram buscados, em Cardiologia e Radiologia, os compostos estudados por Zilio (2009) para que se pudesse observar o seu comportamento em contexto. Os compostos, assim como suas ocorrências, podem ser observados na tabela 2.

Tabela 2. Frequência absoluta dos compostos nominais nos *corpora* de Cardiologia e Radiologia.

| Composto Nominal | Cardiologia | Radiologia |
|---|---|---|
| **Herzinsuffizienz** | 56 | **-** |
| **Risikofaktoren** | 27 | **-** |
| **Zeitpunkt** | 26 | **14** |
| **Herztransplantation** | 16 | **-** |
| **Sinusrhythmus** | 14 | **-** |
| **Herzerkrankung** | 12 | **2** |
| **Herzkatheteruntersuchung** | 7 | **-** |
| **Kontrollgruppe** | 5 | **6** |
| **Ejektionsfraktion** | 5 | **1** |
| **Leistungsfähigkeit** | 5 | **-** |
| **Koronarangiographie** | 1 | **1** |

Como se pode ver, a presença dos compostos se deu de forma muito díspare, com presença de 100% no *corpus* de Cardiologia e somente 44% no de Radiologia. Chamou também atenção o fato de que compostos de frequência alta (cf. Zilio, 2009) como *Herzinsuffizienz* [insuficiência cardíaca] e *Risikofaktoren* [fatores de risco], não apareceram no *corpus* de Radiologia.

Quando se observaram as associações sintáticas dos compostos, percebeu-se que somente *Zeitpunkt* [momento] apresentou um número mais alto de associações parecidas nos dois *corpora*. As associações eram *zum Zeitpunkt der(des)* [no momento da(do)] e *zu(m) (Adjektiv) Zeitpunkt* [em(no) (adjetivo) momento]. Porém, nas ocorrências que não apresentavam essas associações, enquanto *Zeitpunkt*, em Cardiologia, parecia apontar para um momento indeterminado, em Radiologia, alguns

contextos apontaram para um momento específico em uma linha do tempo, ao ponto de o composto ser acompanhado por numerais.

Entre os outros compostos, foi encontrada a associação *Bestimmung der (Adjektiv) Ejektionsfraktion* [cálculo da fração de ejeção (adjetivo)] nos dois *corpora* e, nesse caso isolado, os contextos eram muito próximos. O título do texto de Radiologia de onde provinha a associação esclareceu essa semelhança, pois se tratava de um artigo que abordava tomografia cardíaca. Não por acaso, foi também nesse artigo que se encontravam as únicas duas ocorrências de *Herzerkrankung* [doença cardíaca] e a única ocorrência de *Koronarangiographie*.

Pode-se dizer que, se não houvesse no *corpus* de Radiologia um artigo que se ocupasse do coração, somente dois dos compostos estudados seriam comuns aos dois *corpora*: *Zeitpunkt* e *Kontrollgruppe* [grupo controle]. Podemos, assim, perceber que, em uma observação microestrutural, é claro que se encontrará alguma mistura entre as subáreas, porém, se consegue ver, através do léxico empregado, que estão postos dois tipos textuais diferentes.

As observações interlinguais de textos de Cardiologia mostraram resultados ainda mais interessantes. Novamente se observaram as listas de palavras. Primeiro as listas das 10 palavras mais frequentes (tabela 3).

Tabela 3. As 10 palavras mais frequentes.

| **Alemão** | | | **Português** | | |
|---|---|---|---|---|---|
| **Posição na lista** | Palavra | Frequência relativa (%) | Posição na lista | Palavra | **Frequência relativa (%)** |
| **1** | Der | 3,27 | 1 | de | **5,82** |
| **2** | Die | 2,93 | 2 | e | **2,96** |
| **3** | Und | 2,33 | 3 | a | **2,87** |
| **4** | Mit | 1,5 | 4 | com | **1,66** |
| **5** | In | 1,45 | 5 | o | **1,44** |
| **6** | Patienten | 1,41 | 6 | em | **1,43** |
| **7** | Bei | 1,4 | 7 | pacientes | **1,32** |
| **8** | Eine | 1,15 | 8 | da | **1,18** |
| **9** | Von | 0,96 | 9 | do | **1,12** |
| **10** | Des | 0,75 | 10 | os | **1** |

Ao analisar o contraste das palavras mais recorrentes em Cardiologia, em alemão e português, observou-se que muitas palavras gramaticais “equivalentes” nas duas línguas encontram-se em posições de frequência semelhantes. Esse é o caso de *der*, *des*, *von*, em alemão, e *de*, *da*, *do*, em português, artigos e preposições que indicam, em geral, uma presença grande de genitivo. Outros casos são os de *mit* e *com*; *in*, *bei* e *em*; *und* e *e*. Essa mesma proximidade não se percebeu, de forma tão marcante, na comparação entre com a linguagem comum. Esse dado aponta para uma aproximação possivelmente causada pelo tipo textual, vinculado à linguagem especializada em questão.

Chama a atenção o fato de que o item lexical mais frequente das duas línguas é *pacientes*/*Patienten*, ambos com alta frequência e em posições semelhantes na lista de palavras. Também é interessante o fato de que, tanto em alemão quanto em português, o termo aparece no plural.

Mas esse estudo não observou somente semelhanças. Ao se contrastarem os substantivos mais frequentes (tabela 4), diferenças começam a surgir.

Tabela 4. Os 10 substantivos mais frequentes.

| **Alemão** | | | **Português** | | |
|---|---|---|---|---|---|
| **Posição na lista** | Substantivo | Frequência relativa (%) | Posição na lista | Substantivo | **Frequência relativa (%)** |
| **6** | Patienten | 1,41 | 7 | pacientes | **1,32** |
| **34** | Hypertonie | 0,29 | 25 | tratamento | **0,36** |
| **42** | Gruppe | 0,21 | 27 | estudo | **0,35** |
| **47** | Daten | 0,19 | 29 | doença | **0,33** |
| **49** | Therapie | 0,19 | 33 | grupo | **0,28** |
| **52** | Ergebnisse | 0,18 | 37 | stents | **0,26** |
| **53** | Jahre | 0,17 | 39 | resultados | **0,25** |
| **75** | Transplantation | 0,12 | 43 | cirurgia | **0,22** |
| **86** | Mitralinsuffizienz | 0,11 | 47 | risco | **0,21** |
| **90** | Herzen | 0,10 | 50 | grupos | **0,20** |

No português, os substantivos ocupam posições de destaque na lista, mostrando uma frequência muito mais alta do que no alemão. É importante ressaltar também que, apesar de os textos que

compuseram o *corpus* terem sido selecionados em pares temáticos, os substantivos em comum na lista são somente aqueles que ocorrem em vários textos do *corpus* (*Patienten* - *pacientes*, *Gruppe/Gruppen* - *grupo*, *Therapie* - *tratamento*, *Ergebnisse* - *resultados*). No caso dos outros substantivos da lista, o português e o alemão seguem princípios distintos. No alemão, por exemplo, *Hypertonie* [hipertensão] e *Mitralinsuffizienz* [insuficiência mitral] têm todas suas ocorrências em apenas um artigo. Algo parecido ocorre com *Transplantation* [transplante]. Já no português, *stens* é o único substantivo da lista que ocorre somente em um texto do *corpus*, enquanto a tendência geral é ocorrer em vários artigos.

A partir dessa observação, pôde-se inferir que o vocabulário geral da Cardiologia é muito mais empregado no português, com repetições mais frequentes. De outro lado, em alemão, há uma preferência por se empregar mais o vocabulário específico e por não se repetir muito o vocabulário mais genérico.

Seguindo esse tipo de observação, também se notou uma riqueza lexical maior no alemão, com um índice de 19,65, enquanto no português, o índice atingiu apenas 13,03. Isso indica que o português tende a exibir um vocabulário mais restrito e mais repetitivo do que o alemão[9].

Observando o tamanho das sentenças, o português apresentou sentenças muito maiores, com uma média de 29,23 palavras cada. Já o alemão teve, em média, 20,79 palavras por sentença. Esse é um indício de maior complexidade frasal dos textos em português em relação aos textos em alemão.

Assim como na observação interlingual, foram observados também contextos de termos recorrentes em Cardiologia[10], nas duas línguas. São eles: *Herzinsuffizienz* / insuficiência cardíaca e *Risikofaktoren* / fatores de risco. A associação mais recorrente para o termo *insuficiência cardíaca* foi o adjetivo “congestiva” (*insuficiência*

---

[9] Esse dado, porém, não deve ser tomado como uma certeza, visto que o alemão é uma língua com casos morfológicos, o que influencia no número de possibilidades de flexão de cada palavra, principalmente no caso dos adjetivos. Para confirmar esse indício, seria preciso aplicar um lematizador ao *corpus*.

[10] A recorrência desses compostos em artigos científicos de Cardiologia foi averiguada em Zilio (2009), em um corpus de mais de 2,8 milhões de palavras.

*cardíaca* ***congestiva***). Já o termo *Herzinsuffizienz* é, na maioria das ocorrências, acompanhado pelo adjetivo *schwer* [grave] (*schwere Herzinsuffizienz* [insuficiência cardíaca grave]).

Já com o termo *Risikofaktoren*, a associação mais recorrente foi o adjetivo *unabhängig* (*unabhängige Risikofaktoren* [fatores de risco independentes]). É valido ressaltar que, além da busca pelo termo no plural (*Risikofaktor**en***), foi feita a busca pelo termo na forma singular (*Risikofaktor*). Ainda assim, a associação mais recorrente foi o adjetivo *unabhängig* e ambas as formas tiveram semelhante número de ocorrências (nove no plural e sete no singular). Com o termo em português, por outro lado, não se obteve o mesmo resultado: *fator de risco* teve apenas 7 ocorrências e *fatores de risco*, por outro lado, teve 27 ocorrências. Além dessa diferença em relação aos termos em alemão, não foram encontradas associações recorrentes para os termos em português; cada ocorrência encontrava-se em contextos distintos.

Essas observações permitiram mostrar ao tradutor iniciante que é necessário estar sempre atento ao traduzir textos de Cardiologia do alemão para o português, já há indicações de que o comportamento vocabular seja bastante diferente.

### *4.2 Colocações especializadas*

Este segundo estudo foi desenvolvido em pesquisa de mestrado, por Zilio (2009) e apresentou uma série de colocações especializadas em contraste interlinguístico alemão-português.

O objetivo do estudo foi apresentar para tradutores iniciantes uma série de associações léxico-gramaticais que pudesse facilitar o trabalho de traduzir artigos de Cardiologia do alemão para o português.

Para atingir tal objetivo, tratou-se, primeiramente, de compilar um *corpus* de tamanho razoável (cerca de 2,8 milhões de palavras), dividido em dois subcorpora: alemão e português. Posteriormente, definiu-se o que seria entendido como **colocação especializada**. Foi nesse passo em que a parceria entre a Terminologia Textual e a Linguística de *Corpus* ficou mais óbvia.

Como não havia mais a predeterminação de que o objeto de estudo deveria ser sempre observado a partir de termos, utilizou-se a lista de palavras do subcorpus alemão e, a partir dela, foram

selecionados onze compostos nominais, procurando uma abrangência grande nas ocorrências. Dessa forma, foram selecionados desde os compostos nominais mais frequentes até aqueles com pouco mais de cem ocorrências. A observação a partir dos compostos nominais se justificou pela potencial dificuldade que eles apresentam para a tradução (cf. Leipnitz, 2005).

Tendo selecionado os compostos nominais selecionados (ver tabela 5), buscaram-se os verbos, substantivos e adjetivos deverbais que os acompanhavam. Para realizar essa tarefa, lançou-se mão da ferramenta *Concord*, do software *Wordsmith Tools*. Com essa ferramenta, foi possível verificar cada contexto do *corpus* em que ocorriam os compostos nominais selecionados, além de analisar suas presenças nas formas singular e plural.

Tabela 5. Compostos nominais selecionados com as respectivas frequências, averiguadas na lista de palavras do subcorpus alemão.

| Herzinsuffizienz | 914 |
|---|---|
| Risikofaktoren | 609 |
| Zeitpunkt | 439 |
| Herzerkrankung | 378 |
| Kontrollgruppe | 349 |
| Herztransplantation | 326 |
| Ejektionsfraktion | 221 |
| Koronarangiographie | 220 |
| Sinusrhythmus | 216 |
| Leistungsfähigkeit | 213 |
| Herzkatheteruntersuchung | 101 |

Faltava então determinar o que seria considerado uma colocação especializada e o que seria reconhecido como ocorrência comum. Para tal, utilizou-se um critério de frequência e distribuição bastante abrangente e um critério morfológico. Da seguinte forma: seriam trazidas para o contraste com o português todas as estruturas que apresentassem pelo menos 2 ocorrências em, no mínimo, 2 textos diferentes; sendo que a frequência dessas estruturas era avaliada através da forma lemática. Por forma lemática entendeu-se que

*Durchführung* [realização] e *durchführen* [realizar] pertenciam a um mesmo lema. Assim, uma ocorrência de *Durchführung* [realização] e outra de *durchführen* [realizar] junto a um mesmo composto em textos diferentes já eram suficientes para levar as duas estruturas para um contraste com o português.

Esse contraste também foi feito a partir da ferramenta *Concord*, já que, após terem sido estabelecidos os equivalentes em português para os compostos (ver tabela 6), buscaram-se os contextos desses equivalentes e pôde-se comparar as duas línguas para o estabelecimento de equivalências fraseológicas. Um exemplo é o seguinte:

***an Herzinsuffizienz leiden*** ***vs. insuficiência cardíaca acomete***

| | |
|---|---|
| Weltweit **leiden** mehr als 22 Mio. Menschen **an einer Herzinsuffizienz**. | A **insuficiência cardíaca** é uma síndrome clínica complexa, **acometendo** aproximadamente 4,7 milhões de pessoas nos Estados Unidos. |

Tabela 6. Compostos nominais em alemão e seus equivalentes em português

| Herzinsuffizienz | Insuficiência cardíaca |
|---|---|
| Risikofaktoren | Fatores de risco |
| Zeitpunkt | Momento – Tempo |
| Herzerkrankung | Doença cardíaca - Doença arterial coronariana - Doença isquêmica do coração |
| Kontrollgruppe | Grupo controle |
| Herztransplantation | Transplante cardíaco |
| Ejektionsfraktion | Fração de ejeção |
| Koronarangiographie | Angiografia coronariana |
| Sinusrhythmus | Ritmo sinusal |
| Leistungsfähigkeit | Capacidade funcional - Função cardíaca - Tolerância ao esforço |
| Herzkatheteruntersuchung | Cateterismo cardíaco |

Através desse estudo, foram encontrados 94 pares de colocações especializadas e equivalentes, totalizando 190 estruturas fraseológicas pareadas, contando com centenas de contextos observados.

Esse trabalho foi muito satisfatório, já que, além de permitir a observação de equivalências, também dá acesso aos contextos de ocorrência das expressões levantadas.

## 5. Considerações finais

Neste capítulo, procuramos mostrar, através de exemplos concretos extraídos de trabalhos já realizados, como a Terminologia Textual e Linguística de *Corpus* são áreas que trabalham muito bem em parceria.

Principalmente por serem áreas de estudo interdisciplinares, elas se colocam muito bem à disposição do linguista para interagir não só entre si, mas com muitas outras áreas. Nos estudos que apresentamos, por exemplo, pudemos observar como ambas interagem com os Estudos de Gêneros Textuais, com a Linguística Contrastiva, com os Estudos Fraseológicos, com os Estudos da Tradução etc.

Também cremos ter ficado claro como os estudos não se dão somente em um nível mais superficial, mas possibilitam que se observe o comportamento linguístico em contexto, permitindo ao linguista olhar seus dados a fundo.

É preciso que se reconheça, porém, que tanto a Terminologia Textual quanto a Linguística de *Corpus* ainda se apresentam em estágio de desenvolvimento, não tendo recebido grande atenção em nosso país até o momento. Pode-se perceber, porém, pelos estudos já citados e por muitos outros estudos (Hoffmann, 1998b, 1998c; Finatto et alii, 2003; Finatto, 2004b; Azeredo, 2007; etc.), que essas áreas contribuem muito e podem contribuir ainda mais para o desenvolvimento de nossa ciência linguística.

Existem naturalmente alguns questionamentos e posicionamentos que ainda se mantêm sem solução definitiva, como foi comentado neste artigo. Sendo um dos principais deles a decisão sobre a situação da Terminologia Textual: deve ela ser entendida como um tipo

de Terminologia ou talvez como uma área que engloba a Terminologia (como uma Linguística de Textos Especializados ou de Linguagens Especializadas)?

Para esclarecer esses questionamentos, cremos que o crescimento dessas áreas de estudo e a sua maior aceitação entre a comunidade de linguistas é a melhor solução. De forma que uma maior visibilidade dessas áreas de estudo e uma discussão em nível nacional deve ser o melhor caminho para o seu assentamento no quadro teórico da Linguística brasileira.

## Referências bibliográficas

BERBER SARDINHA, T. (2004) *Linguística de Corpus*. Barueri: Manole.

BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus Linguistics*: Investigating language structure and use. Cambridge: Cambridge University Press, 1998.

CABRÉ, M.T. (2001) *La Terminologia.* Representación y comunicación. Barcelona: IULA.

CIAPUSCIO, G. E. (1994) *Tipos textuales*. Buenos Aires: Universidade de Buenos Aires.

CIAPUSCIO, G. E. (2003) *Textos especializados y terminología*. Barcelon: UPF/IULA.

CONDAMINES, Anne. (2005) Linguistique de corpus et terminologie. In: *Langages*, 157, p. 36-47. Disponível em: http://w3.univ-tlse2.fr:8880/erss/index.jsp?perso=acondami&subURL=Langagesdef.pdf. Acessado em 10/06/2008.

FINATTO, M.J.B.; EICHLER, M.L.; DEL PINO, J.C. (2003) Sujeitos e agentes de poder e dever em textos sobre equilíbrio químico: aspectos lingüístico-terminológicos e aspectos conceituais da enunciação científica e o ensino-aprendizagem de Química. In: *Organon*, 32-33, vol.16, p.83-104.

FINATTO, M.J.B. (2004a) Termos, textos e textos com termos: novos enfoques dos estudos terminológicos de perspectiva linguística. In: ISQUERDO, A.N.; KRIEGER,M.G. (Orgs.). *As ciências do léxico*: lexicologia, lexicografia, terminologia, volume II. Campo Grande: Ed. UFMS.

FINATTO, M.J.B. (2004b) Terminologia e Linguística de Corpus: da Perspectiva Enunciativa aos Novos Enfoques do Texto Técnico-científico. In: *Letras de Hoje*, Porto Alegre, v.39, n.4, dezembro, p.97-106.

HOFFMANN, L. (1988a) *Vom Fachwort zum Fachtext*: Beitrage zur Angewandten Linguistik. Tübingen: Gunter Narr Verlag.

HOFFMANN, L. (1988b) Grundbegriffe der Fachsprachenlinguistik. In: *Germanistisches Jahrbuch für Nordeuropa*. 7. Folge. Deutsche Fachsprachen in Forschung und Lehre. Helsinki, Estocolmo: p. 9-16.

HOFFMANN, L. (1998a) Syntaktische und morphologische Eigenschaften von Fachsprachen. In: HOFFMANN, L. et al. (Orgs.) *Fachsprachen: ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. Walter de Gruyter, p. 416-427.

HOFFMANN, L. (1998b) Anwendungsmöglichkeiten und bisherige Anwendung von statistischen Methoden in der Fachsprachenforschung. In: HOFFMANN, Lothar; KÄLVERKÄMPER, Hartwig; WIEGAND, Herbert Ernst (Orgs.). *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. Berlin, New York: Walter de Gruyter, p. 241-249.

HOFFMANN, L. (1998c) Anwendungsmöglichkeiten und bisherige Anwendung von linguistischen Methoden in der Fachsprachenforschung. In: HOFFMANN, Lothar; KÄLVERKÄMPER, Hartwig; WIEGAND, Herbert Ernst (Orgs.). *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. Berlin, New York: Walter de Gruyter, p. 241-249.

KALVERKÄMPER, H. (1983) Textuelle Fachsprachen-Linguistik als Aufgabe. In: *Zeitschrift für Literaturwissenschaft und Linguistik*, v. 51/52, n° 13, p. 124-166.

KRIEGER, M.G. (2008) Porque Lexicografia e Terminologia: relações textuais? In: FINGER, I. & COLLISCHONN, G. (Orgs.) *Anais do 8° Encontro do Círculo de Estudos Lingüísticos do Sul - CELSUL*, Porto Alegre, de 29 a 31 de outubro de 2008. Pelotas: EDUCAT. ISBN: 978-85-7590-115-1.

LEIPNITZ, L. (2005) Compostos nominais em língua alemã em medicina em tradução para o português. Porto Alegre: UFRGS – PPG-Letras – Estudos da Linguagem. Dissertação de mestrado. Linha de pesquisa: Lexicografia e Terminologia: Relações Textuais. Orientadora:

Maria José Bocorny Finatto.

LEITCHIK, V.M. (2002) Elements of Terminological Text Theory. In: SHELOV, S.D.; LEITCHIK, V.M. (eds.) *Russian Terminology Science (1992-2002)*. Viena: Termnet Publisher. Editado com a colaboração de H. Picht e C. Galinski.

RAJAGOPALAN, K. (2007) A Linguística de Corpus no tempo e no espaço: visão reflexiva. In: GERBER, R.M.; VASILÉVSKI, V. *Um percurso para pesquisas com base em corpus*. Florianópolis: Editora da UFSC.

SCHEEREN, F.; MIGOTTO, E. J.; ZILIO, L. (2008) Estudo Exploratório sobre artigos de Cardiologia em Alemão e Português: Macroestruturas e usos dos termos Herzinsuffizienz/insuficiência cardíaca. In: XX Salão de Iniciação Científica, 2008, Porto Alegre. *CD - Resumos do XX Salão de Iniciação Científica*. Porto Alegre: UFRGS.

SINCLAIR, J.M. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

TEMMERMANN, R. (2000) *Towards New Ways of Terminology Description*. The sociocognitive-approach. Amsterdam, Philadelphia: John Benjamins Pub. Co.

WÜSTER, E. (1974) Die Allgemeine Terminologielehre. Ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften. In: *Linguistics*, v. 119, Janeiro de 1974, p. 61-106.

ZILIO, L. (2009) *Colocações especializadas e Komposita: um estudo contrastivo alemão-português na área de cardiologia*. Porto Alegre: UFRGS. Dissertação de Mestrado. PPG-LETRAS/UFRGS.

ZILIO, L.; SCHEEREN, F.; FINATTO, M.J.B. (2009) Artigos científicos de Cardiologia: contraste de macro e microestruturas para caracterização de tipo textual. In: SIGET, Simpósio Internacional de Estudos de Gêneros Textuais, 5, 11 a 14 de agosto de 2009, Caxias do Sul, RS, Brasil. *Anais do 5° Simpósio Internacional de Estudos de Gêneros Textuais*. Caxias do Sul: UCS. *(no prelo)*

# PARA ALÉM DAS TERMINOLOGIAS: ESTUDOS DE CONVENCIONALIDADE EM LINGUAGENS CIENTÍFICAS

*Maria José Bocorny Finatto*

*Aline Evers*

*Cybele Margareth de Oliveira Alle*

***Resumo**: Este capítulo relata algumas explorações sobre padrões de convencionalidade em linguagens científicas. São examinados conectores causais em corpora de Química e de Pediatria, com destaque para PORQUE e BECAUSE em artigos de revistas brasileiras e internacionais de Pediatria. A apresentação de dados é ilustrativa e não há pretensão de rigor estatístico. As investigações foram feitas à luz dos enfoques da Linguística de Corpus e de um tipo de estudo sobre linguagens e textos científicos que não toma terminologias como elementos de partida. São apresentados os enfoques teórico-metodológicos utilizados e situada a causalidade como um tópico de investigação. Em seguida, é apresentado um quadro geral da exploração empreendida, na qual foram verificadas as seguintes tendências: a) usos diferenciados de conectores por áreas de conhecimento; b) BECAUSE como forma mais presente entre os conectores causais verificados em revistas internacionais, enquanto há pouco uso de PORQUE nas revistas brasileiras; c) a transferência da variedade de conectores causais dos artigos brasileiros de Pediatria para suas versões em inglês parece gerar um padrão de inglês distinto do inglês internacional. Por fim, o trabalho evidencia o bom potencial de um enfoque mais global para as linguagens especializadas articulado à Linguística de Corpus e valida o exame de conectores causais para o reconhecimento de especificidades dos gênero textuais das linguagens científicas.*

## 1. Introdução

Este trabalho apresenta dados obtidos em uma série de pesquisas linguísticas exploratórias com *corpora*. O interesse das explorações foi

identificar prováveis padrões preferenciais para expressar causa em textos científicos brasileiros. Essas pesquisas integraram o Projeto *Causalidade no texto de Química: coesão, terminologias e enunciação científica*, desenvolvido com apoio do CNPq[1].

Os materiais estudados pertencem a dois gêneros textuais (cf. Swales, 1990): o primeiro gênero é o texto didático universitário; o segundo é o artigo científico publicado em revistas especializadas. Como *corpora*, utilizamos quatro grupos de textos:

a) um conjunto de textos que correspondem a capítulos completos de livros didáticos de Química Geral, Física, Química e Pediatria largamente utilizados em universidades brasileiras;

b) um conjunto de artigos de revistas brasileiras de Química e de Física produzidas com o aval de associações nacionais de pesquisa;

c) um conjunto de textos originais em português e suas respectivas versões para o inglês publicados em uma prestigiada revista bilíngue brasileira de Pediatria, editada pela associação nacional dessa especialidade;

d) um conjunto de textos originais em inglês publicados em duas revistas norte-americanas de Pediatria que têm estatuto internacional.

Este texto, além de dar a conhecer resultados da sua parte experimental, também tem o interesse evidenciar a validade de aproximar dois enfoques teórico-metodológicos, da Linguística de *Corpus* e o da *Linguística das Linguagens Especializadas*. Essa aproximação, conforme acreditamos, pode gerar subsídios interessantes para a qualificação do ensino de tradução em áreas científicas, tais como Química e Medicina/Pediatria.

## 2. Dos enfoques envolvidos

O primeiro enfoque que guiou nossas explorações sobre conectores causais em textos científicos tem um prestígio inegável no cenário brasileiro (Rajagopalan, 2007). Esse enfoque advém da Linguística de *Corpus* (doravante **LC**).

---

[1] Mediante Bolsa PQ, Produtividade, processo 301102/2006-6, vigência até 28/02/2010.

A **LC** será aqui caracterizada, muito resumidamente, em seus princípios teóricos mais elementares e procedimentos de abordagem estatística de usos, de combinatórias e de repetição de palavras em *corpora*. Nossas fontes de referência são os trabalhos de Berber Sardinha (2004), precursor da **LC** no cenário brasileiro, os de Stubbs (1996 e 2001) e os de Sinclair (1991).

Voltaremos à **LC** mais adiante, visto que o segundo enfoque a apresentar – a *Linguística das Linguagens Especializadas* – necessita de uma contextualização um pouco mais detida, dado que é bem menos conhecido no Brasil.

## *2.1. Linguística das Linguagens Especializadas*

Essa perspectiva corresponde, para os fins deste trabalho, à reunião de ideias e de propostas de estudo ou de composição de programas teóricos identificadas genericamente pelos nomes de *Terminologia Textual*, *Terminologia Textualista*, *Teoria Terminológica do Texto*, *Linguística das Linguagens Especializadas* ou *Linguística do Texto Especializado*.

Esses estudos, cuja denominação é heterogênea, têm em comum o emprego de metodologias de investigação centradas na descrição macro e microestrutural de conjuntos de textos de modo extensivo e com apoio estatístico (Hoffmann, 2004 e outros anos; Kocoureck, 1991, Leitchik, 2004). Outro traço partilhado entre as suas diferentes tendências é a função importante que tem a verificação de um *modus dicendi* técnico-científico.

Seus propósitos são descritivo-explicativos e compartilham também a opção por colocarem as terminologias em meio a todo um conjunto de fenômenos da comunicação técnico-científica (Finatto, 2007(a), 2007(b) e 2008), sem que, no entanto, essas terminologias sejam os objetos centrais ou privilegiados das investigações.

De um modo geral, em meio a um ambiente textual sujeito às convenções dos gêneros textuais implicados, torna-se central todo um conjunto de elementos sintáticos, semânticos e pragmáticos que constituem as práticas textuais técnico-científicas.

Esse enfoque não despreza as terminologias, vale sublinhar. Apenas não as toma como centro gravitacional, visto que o macro-plano do texto assume tal estatuto. A investigação, assim, torna-se mais ampla, abordando-se os usos das linguagens técnicas ou científicas, tal como realizadas em textos, orais ou escritos, incluindo as re-elaborações das linguagens de divulgação científica ou tecnológica para leigos.

Outra característica da perspectiva ampliada da *Linguística das Linguagens Especializadas* é a relevância atribuída à articulação entre a significação da forma ou do modo de dizer do texto e o seu significado nocional. O texto é visto como um todo de sentido e de comunicação, constituído pelo conteúdo expresso e também pelo quadro geral das escolhas ou convenções do seu "modo de dizer" estabelecidos pela comunidades discursivas e interpretantes envolvidas. A Figura 1 abaixo visa representar essa articulação entre o que é dito/posto no texto e as escolhas realizadas para a sua formulação. Desenha-se uma associação entre enunciados e enunciações conforme já desenvolvido na perspectiva benvenistiana de uma *Linguística da Enunciação* (BENVENISTE, 1989 e outros):

Figura 1: Concepção de texto no enfoque global da *Linguística das Linguagens Especializadas*.

Nesse entendimento de texto, não se deixam de acolher, com igual importância dada às terminologias, elementos tais como conjunções, enunciados definitórios, construções frasais, combinatórias de palavras mais ou menos fixas, elementos retóricos e

argumentativos, fraseologismos, condições pragmáticas da comunicação, condições do gênero textual, modalizações e outros recursos. O foco de atenção principal recai sobre usos concretos e padrões a partir deles verificáveis.

Essa concepção de trabalho não deve ser confundida com estudos terminológicos *stricto sensu*. Entretanto, justamente por suas diferenças e complementaridades, cremos que há uma boa interface de enriquecimentos e de aprendizagens mútuos entre a *Linguística das Linguagens Especializadas* e os estudos terminológicos.

Hoffmann (1988a), por exemplo, já cultivava a ideia de um enfoque global da linguagem científica ou técnica quando apresentou-nos seu projeto de uma *Linguística de Linguagens Especializadas*. Essa *Linguística* Kalverkämper (1983, p. 125) já descrevia como uma "**teoria terminológica com uma série de problemáticas adicionais**"[2] (grifo nosso).

Desde o início da década de 80, essa perspectiva de estudo das Linguagens Especializadas coloca-se como algo que deveria abranger mais do que o léxico, ainda que o componente lexical tenha sido considerado frequentemente como o "*Hauptträger der Fachsprachlichkeit*" [principal portador da especialização linguística] (cf. Kalverkämper, 1983, p. 125). Sua ideia-base é que essa especificidade não está somente no vocabulário.

Leitchik (2004) é um outro autor que integra o grupo dos estudiosos de um enfoque mais global das linguagens especializadas, com uma trajetória de reflexão que também iniciou nos anos 80. Em uma coletânea de artigos de autores russos sobre Terminologia do período de 1992 a 2002, apresenta uma proposta de trabalho que denomina de *Teoria Terminológica do Texto* [*Terminological Text Theory*], a qual já vinha desenvolvendo desde 1981 (cf. Leitchik, 2004, p. 373). Não se afasta radicalmente do conceito de termo e de sua importância para os estudos terminológicos, porém, reconhece que o *objeto de estudo* dessa sua *Teoria Terminológica do Texto* compreende vários textos, nos quais termos estão contidos (Leitchik, 2004, p. 371).

---

[2] No original: "Terminologielehre mit einem zusätzlichen Fächer von Problemstellungen".

Aponta também que a sua desejada Teoria "se ocupa do conceito de texto propriamente dito e, dessa forma, também com os conceitos de outras informações textuais (categorias, qualidades) necessárias: coesão, completude, "separatividade", significado emotivo (direção comunicativa)". Essa perspectiva parece afinar-se com a ideia de trabalho com *corpora* à medida que reconhece o papel de *vários textos*.

Ale, desses dois autores, cumpre registrar a ideia original da "*Fachtextlinguistik*"[3] (cf. Kalverkämper, 1983). Esse termo, em alemão, conforme a ênfase, poderia ser interpretado de duas maneiras: primeiro, temos uma "*Fach-Textlinguistik*" [Linguística do Texto em Especialidades]. Essa *linguística* buscaria responder o que é uma especialidade, ocupando-se mais especificamente da pragmática, "pois a busca pela especialidade ou especialidades conduz às ações humanas"[4] (Kalverkämper, 1983, p. 128).

Na segunda possibilidade de entendimento do termo *Fachtextlinguistik*, temos uma "*Fachtext-Linguistik*" (Linguística do Texto Especializado), entendida como uma *Linguística de gêneros textuais* e que se ocupa dos textos em funcionamento.

Utilizando o conceito de gênero textual como base, como aponta Kalverkämper (1983, p. 145), "a *Linguística do Texto Especializado* poderia se apoiar também na Linguística das variedades diassistemáticas, que observa e analisa os diferentes 'níveis', 'registros' e 'variedades' como subsistemas do sistema linguístico"[5].

---

[3] *Fachtextlinguistik*, em alemão, se não tiver nenhum sinal gráfico que aponte qual a divisão que se deve fazer na palavra, pode ser interpretada de duas formas. A primeira forma *Fach-Textlinguistik*, seria uma Linguística Textual Especializada; já a segunda forma possível, *Fachtext-Linguistik*, seria uma Linguística do Texto Especializado. A diferença entre as duas formas é mais aprofundada em Kalverkämper (1983). Neste estudo, ela é abordada somente para apontar as diferenças mais básicas.

[4] No original: "denn die Frage nach dem Fach oder den Fächern zielt auf das menschliche Handeln".

[5] No original: "die Fachtext-Linguistik könnte sich auch (…) an der Varietätenlinguistik [orientieren], die verschiedenen, Niveaus'/,Register'/,Varietäten' als Subsysteme des Sprachsystems beachtet und analysiert". Nota da tradução: a opção de tradução por **Linguística de variedades diassistemáticas** considerou a menção pelo autor às propostas de Eugênio Coseriu, um importante linguista que propunha que um sistema de

Esse último entendimento do autor que funda essa perspectiva, acreditamos, resume em boa medida um projeto epistemológico comum. Ao mesmo tempo, pode nos oferecer um perfil de identidade do enfoque mais global sobre linguagens especializadas. É esse projeto comum que identificamos aqui como *Linguística das Linguagens Especializadas*. Vai além do texto, pois abrange vários textos e gêneros textuais, e questiona-se sobre a conformação da linguagem empregada nas situações de comunicação científica, tecnológica e profissional. Dito isso, voltemos a tratar da LC.

## *2.2 Linguística de Corpus*

A **LC** pode ser definida não como uma nova Linguística, mas como um novo caminho para a Linguística (Berber Sardinha, 2004, p. 10). Mostra-se como uma metodologia e também como uma abordagem diferenciada dos estudos da linguagem que se distingue, fundamentalmente, dos estudos de herança gerativista.

A **LC** é marcada pela observação, o mais extensiva possível, dos usos reais da língua e pelo apoio de recursos informatizados aplicados a acervos textuais em formato digital criteriosamente organizados. Esses acervos são os *corpora*.

Nesse enfoque, o importante é reconhecer, nos *corpora*, a variedade extensiva dos desempenhos linguísticos com vistas à depreensão *a posteriori* de padrões. As fontes de dados devem permitir a identificação de padrões e de especificidades dos usos da língua em diferentes situações, inclusive na macro-situação dos textos técnico e científicos.

A língua, sob a ótica da LC, passa a ser entendida como um sistema probabilístico de combinatórias, de modo que uma dada

---

língua é composto por um *continuum* de vários sistemas. Assim, o subsistema da linguagem da Economia em português brasileiro integra o sistema geral da língua portuguesa. Mais tarde, esse tipo de estudo sobre variabilidades seria plenamente desenvolvido em trabalhos que se ocupam da Variação Linguística, tais como os de W. Labov, autor que levou-nos a um alargamento da noção de *competência* proposta por Chomsky ao verificar sistematicidades das variações de língua inglesa manifestadas por determinados grupos sociais, tidas até então apenas como idiossincráticas.

palavra se define pela sua presença e pelos tipos de vinculação com as demais palavras dessa língua ao longo de todo um *continuum* de usos. Assim, para as palavras, vale uma adaptação da máxima "Diga-me com quem andas e te direi quem [o que] és".

Os princípios mais básicos da LC podem ser assim resumidos:

- Visão sociolinguística da comunicação;
- Prioridade dos dados empíricos: não há distinções ou condições especiais para qualquer palavra *a priori*, é preciso que um *corpus* as demonstre;
- Visão probabilística da linguagem;
- Configuração sistêmico-funcional do significado no contexto.

Dois dos principais pressupostos da **LC**, conforme depreendemos das ideias de Stubbs (1996, p. 46) e de Sinclair (1991, p. xviii), são que:

a) um *corpus* não é mera ferramenta de análise, mas um importante conceito teórico;

b) a linguagem se mostra diferente quando examinada extensivamente.

Feita essa brevíssima caracterização da **LC**, passamos agora a apresentar o tópico de estudos da causalidade em meio ao nosso tipo de investigação sobre textos científicos. Tratou-se assim, de uma investigação que apostou e segue apostando na viabilidade de uma cooperação entre **LC** e o enfoque *Linguística das Linguagens Especializadas*.

## 3. Causalidade

Nas linguagens científicas, o reconhecimento do modo como as relações de causa se estabelecem e se expressam é fundamental, uma vez que os textos produzidos, tal como os textos de Química Geral ou os artigos de Pediatria de que trataremos a seguir, pretendem informar, atualizar, propagar e ampliar conhecimentos científicos.

Considerando os propósitos comunicativos dos textos científicos em suas diferentes modalidades, cremos que encadeamentos claros, tal como os de coesão e coerências textuais, por exemplo, são fatores importantes para que sejam compreendidas principalmente as causas e

os efeitos de procedimentos descritos, testados e validados. Portanto, a causalidade é um tópico relevante para quem se procure descrever o funcionamento das linguagem científicas.

Em que pese sua importância, a causalidade, como bem se sabe, é tema vasto e complexo, podendo ser estudado sob inúmeras perspectivas diferentes. No âmbito dos estudos da linguagem, o tratamento da causalidade geralmente é bastante marcado pela descrição de uma lógica proposicional associdada a construções lógico-semânticas. E há, também, estudos linguísticos que apontam que haveria uma causalidade do enunciado e outra da enunciação.

Na perspectiva dos estudos da linguagem, como assinalam Neves, Souza e Domingos (1999, p. 493), o estudo das construções causais ocorrentes em uma língua natural não se pode governar apenas pela verificação de esquemas lógicos estritos. Segundo as autoras, um estudo sobre modos de expressar causa terá que, necessariamente, abrigar a investigação do complexo de domínios envolvidos na produção dos enunciados causais.

A concepção de causalidade que utilizamos em nossas explorações de *corpora* assentou-se sobre as ideias de Neves (2001), autora que nos propõe um *continuum* de funcionamentos causais para uma série de conectores. Essa sua ideia, ampliamos também, posteriormente, para evrbos causais.

Entre os conectores causais identificados pela autora, estão alguns cujo sentido canônico, literal ou mais prototípico, não é de causa, tal como, por exemplo, PORTANTO ou ASSIM. Nessa direção de heterogeneidades de sentido, a autora chama atenção para o fato de que a noção de casualidade não corresponde a um sentido preciso e bem delimitado. Antes disso, trata-se de sentidos que podem englobar elementos tidos pelos estudos de tradição gramatical como elementos conclusivos, consecutivos, explicativos ou conformativos.

Na perspectiva de uma “funcionalista em contexto”, Neves assume justamente a importância da verificação contextual de **funcionamentos causais (grifos nossos)** para que se possam produzir descrições do funcionamento da língua portuguesa. Desse modo, um conector X ou um verbo Y não é ou será necessariamente causal: ele, antes disso, estaria em usos mais ou menos causais.

Entre todo um universo de encadeadores lógicos e textuais e de verbos que exprimem causa, nossa atenção recairá aqui apenas sobre os conectores frasais e oracionais reconhecidos como potencialmente causais por Neves (2001).

A opção por examinar esses elementos deu-se porque, em ocasião anterior, quando estudamos textos de Química (Finatto e Simioni, 2007; Finatto *et al.*, 2006), verificamos que a causalidade se estabelecia, em nosso *corpus*, muito mais através do uso de conectores do que através de construções predicativas verbais.

Naquela ocasião, guiados pela concepção de que funcionamentos causais deveriam mostrar-se em contextos de uso, examinamos, em cinco livros didáticos de Química Geral – em um *corpus* que totalizou 429.624 palavras (*tokens*) e apresentou 11.818 palavras diferentes (*types*) –, a presença de 22 conectores que reconhecemos como potencialmente causais. Os conectores foram os seguintes:

1. PORQUE
2. JÁ QUE
3. UMA VEZ QUE
4. DADO QUE
5. DESDE QUE
6. VISTO QUE
7. SENDO QUE
8. LOGO
9. ENTÃO
10. PORTANTO
11. ASSIM
12. POR ISSO
13. ENFIM
14. POR FIM
15. CONSEQUENTEMENTE
16. POIS
17. DEVIDO
18. POR CAUSA D*
19. DE MODO QUE
20. DE MANEIRA QUE
21. DE FORMA QUE
22. POR CONSEGUINTE

Quisemos, assim, verificar quão recorrente seria a presença desses elementos e de que modo sentidos causais mais ou menos literais, dados *a priori,* funcionariam contextualmente no nosso *corpus* de Química. Para localização dos contextos de cada elemento, utilizamos as ferramentas para análise de *corpora* disponíveis no *site* do prof. Tony Berber Sardinha da PUC-SP[6].

A partir de uma primeira exploração, apenas quantitativa, percebemos que a frequência específica de alguns destacava-se naquela área e naquele gênero textual didático. Concluímos que as concentrações de alguns e o rareamento de outros, ao longo de diferentes textos que abordavam diferentes temas de Química Geral, poderiam ser fator indicativo da especificidade do gênero e da linguagem científica sob exame. Naquele quadro, ENFIM, por exemplo, não ocorria em nenhum texto do *corpus*. PORTANTO, de modo inverso, atravessava quase todos os textos com percentuais de uso aproximados em relação ao número de palavras de cada texto.

Então, para ampliar o contraponto com indícios que associavam temas dos textos, gênero textual e diferentes frequências de conectores causais, recorremos a uma nova amostra de textos. A nova amostra foi composta por um manual didático de Físico-Química, artigos de periódicos de Química, das revistas *Química Nova* (v. 27, 2004) e *Química Nova na Escola* (nº 19 e 20 de maio e novembro de 2004, respectivamente).

E, para uma maior amplitude na observação, verificamos também a presença desses conectores de causalidade em textos de revistas de popularização de temas de ciências, os quais têm, em tese, alto grau de didatismo, mas pertencem ao gênero jornalístico. Esse gênero foi representado por matérias sobre Química na revista *Superinteressante* (textos aleatoriamente selecionados, dos anos de 88 a 94). O *corpus* utilizado para contrastar os dados obtidos em Química Geral contabilizou um total de 609.997 *tokens* e 37.027 *types*.

As expressões mais frequentemente utilizadas nos diferentes gêneros textuais foram: ASSIM; PORTANTO; ENTÃO; POIS; PORQUE; DEVIDO E COMO.

---

[6] http://www2.lael.pucsp.br/corpora/index.htm.

O Quadro 1, a seguir, apresenta os itens mais frequentes, em ordem de utilização, tanto no *corpus* de estudo de Química Geral como nos textos dos *corpora* utilizados para contraste.

| **FONTES DE TEXTO** | **CONECTORES MAIS PRESENTES (POR ORDEM DE USO)** |
|---|---|
| Manuais didáticos de Química Geral | ENTÃO – 0,1575%<br>PORTANTO – 0,1575%<br>ASSIM – 0,13%<br>PORQUE – 0,0975% |
| Manuais didáticos de Físico-Química | ENTÃO – 0,170%<br>PORTANTO – 0,152%<br>ASSIM – 0,126%<br>COMO – 0,098% |
| Artigos revista Química Nova | DEVIDO – 0,0583%<br>POIS – 0,0416%<br>ASSIM – 0,035%<br>PORTANTO – 0,0243% |
| Artigos revista Química Nova na Escola | ASSIM – 0,087%<br>PORTANTO – 0,041%<br>POIS – 0,038%<br>DEVIDO – 0,02% |
| Artigos revista Superinteressante | ASSIM – 0,09%<br>PORQUE – 0,08%<br>POIS – 0,07%<br>ENTÃO – 0,06% |

Quadro 1: conectores mais presentes nos *corpora* de Química por tipo de texto. O percentual é a relação, sem ponderação, entre o número de ocorrências do conector e o número de palavras do *corpus*.

Para reavaliar essas tendências, em novembro de 2008, extrapolamos o âmbito da Química. Tomamos, do nosso grupo inicial de 22 conectores com provável sentido de causa, os elementos DEVIDO, ASSIM, POIS, COMO, PORTANTO, LOGO (um item pouco utilizado na exploração anterior), PORQUE e ENTÃO. O objetivo, na nova busca, foi o de verificar se uma concentração semelhantemente variável se daria em outros *corpora* e se haveria alteração nas

preferências e nos usos desses conectores em textos com grau maiores ou menores de didatismo. Foi nossa intenção também verificar tendências com conectores cujo sentido causal tinha interface com um sentido conformativo e/ou conclusivo.

Construímos, assim, um novo *corpus*, composto por cinco artigos de cada uma das seguintes revistas especializadas: *Química Nova na Escola*, 2004 (*tokens* 11.457, *types* 2.874); *Química Nova*, 2004 (*tokens* 19.994, *types* 2.111); *Física na Escola*, 2007 (*tokens* 19.060, *types* 3.448); *Física Aplicada e Instrumentação*, 2005 (*tokens* 18.010, *types* 2.967), e *Jornal de Pediatria* (JPED) nos volumes de 2003 (*tokens* 12.956, *types* 2.345). Nesse *corpus* foram incluídos cinco capítulos (originalmente publicados como artigos) do manual didático universitário *Pediatria: diagnóstico e tratamento*, 2005 (*tokens* 27.136, *types* 4.565).

Foram considerados como textos de maior didatismo os textos das revistas *Química* e *Física na Escola* e os textos do manual didático de Pediatria. Essa avaliação sobre grau de didatismo dos textos foi feita mediante leitura e a consideração de maior acessibilidade de entendimento para leitores não familiarizados com os temas tratados. Os leitores não familiarizados foram as três autoras deste trabalho e mais uma bolsista, linguistas-tradutoras sem qualquer formação em Química, Física ou Ciências da Saúde.

Do conjunto dos textos com 25 artigos e cinco capítulos de um livro, através das ferramentas para exploração de *corpora* disponíveis no *site* www.ufrgs.br/textquim, chegamos a um total de **873** contextos para os conectores causais selecionados para exame. Quisemos, no entanto, não apenas mapear esses conectores e tratar quantitativamente a sua presença. Havia chegado, enfim, o momento do estudo qualitativo sobre o funcionamento do conector com sentido causal.

Verificamos, através de uma leitura de todos os 873 contextos feita por quatro pessoas[7], que tínhamos, entre nós, várias divergências de percepção do sentido de causalidade. Também vimos que essas divergências manifestavam-se mais em contextos de determinados conectores – como os de PORTANTO, por exemplo. Buscamos, então,

---

[7] Além das três autoras deste trabalho, contamos com a participação da acadêmica de Tradução Paula Marcolin.

verificar, munidas de critérios de julgamento objetivos, quais dos conectores listados pela ferramenta geradora de contextos estariam realmente sendo empregados com um sentido de causa. Vejamos abaixo dois contextos examinados:

Importantes correlações entre o universo macroscópico e o universo atômico, tais como a indestrutibilidade dos átomos ao longo de uma reação química, de um lado, e a lei da conservação da matéria de Lavoisier, de outro, não são trazidas à tona. Perde-se, **assim**, uma excelente oportunidade de discutir a própria estrutura metodológica da Química, transitando entre modelos, fatos e sua descrição (Fonte: revista *Química Nova na Escola*)

As misturas foram agitadas por 24 h. Em seguida, filtraram-se as amostras em papel de filtro qualitativo, determinando-se **então** a concentração restante dos íons metálicos nas soluções, por espectrofotometria de absorção atômica. As quantidades dos íons adsorvidos foram determinadas por diferença entre a quantidade adicionada inicialmente e a presente na solução após o tempo de agitação considerado. (Fonte: revista *Química Nova*)

### *3.1 Da verificação de usos com sentido causal*

Para verificar o sentido de causalidade em contextos de ocorrência, seguimos os procedimentos sugeridos por Neves (2000) e Lopes (2001). Esses procedimentos implicaram leitura e aplicação de testes.

Um procedimento sugerido por Neves, que aqui chamamos de teste de *topicalização*[8], consiste em modificar a ordem dos eventos da frase, trazendo a causa para o início da sentença através do uso de uma estrutura que segue o seguinte padrão: É POR CAUSA DE *(causa)... QUE (efeito)* ou *É PORQUE (causa)... QUE (efeito).* Esse procedimento é um dos mecanismos que o enunciador utiliza quando quer enfatizar a causa, movendo-a, portanto, para o começo da sentença.

[8] Esse procedimento é também conhecido como teste de clivagem, embora o que utilizamos na pesquisa não seja o processo de clivagem *stricto sensu*. A autora, ao mencionar a ordem nas construções causais, afirma que quando o enunciador coloca um PORQUE no início da sentença o faz para focalizar a causalidade. Neves (2008, p. 808) chama esse procedimento, entre outros citados, de clivagem (Verbo Ser + Porque).

Encontramos na bibliografia sobre o tema também um outro teste de verificação de sentido de causalidade baseado em procedimento de substituição. No trabalho de Freitas e Ramilo (2004)[9], os conectores sob avaliação são trocados por expressões que ajudem a salientar um sentido de causalidade, tais como POR CAUSA DISSO, POR ISSO, EM FUNÇÃO DISSO.

Aplicamos esses dois tipos de procedimentos para a avaliação de sentido de casualidade. Para ser considerado aprovado o sentido causal do conector, o contexto deveria ser bem-sucedido simultaneamente nos dois testes, topicalização e substituição. Mas, considerando que éramos quatro avaliadoras, estipulamos que a aprovação de uso com sentido causal deveria ser de pelo menos três de nós.

A partir desses dois testes, examinados os 873 contextos obtidos para os conectores DEVIDO, ASSIM, POIS, COMO, PORTANTO, LOGO, PORQUE e ENTÃO. Com a aplicação dos dois testes, chegamos ao número de **309** contextos em que os conectores estavam sendo usados com sentido causal.

O que nos chamou atenção, entretanto, foi a grande massa formada por conectores que, *a priori*, possuíam sentido de causa, mas que, após passarem pelas leituras e pelos testes, geraram uma percepção de causalidade duvidosa. Essa dúvida permanecia, sobretudo, nos casos de reprovação em um dos dois testes ou de desaprovação por duas das quatro examinadoras. Vejamos um dos casos duvidosos:

Apesar do pequeno tamanho desta casuística, a consonância desta distribuição com os resultados relatados por outros investigadores reforça a urgência de se tornar disponível uma vacina que inclua o sorotipo cinco, para ser utilizada na população geral, oferecendo **assim** uma proteção mais ampla. (Fonte: *Jornal de Pediatria*)

Essa flutuação de percepção do sentido apontava talvez para a nossa incapacidade de julgamento e para o fato de nãos sermos especialistas nas áreas tratadas nos textos. Entretanto, verificamos

---

[9] Nosso teste de substituição foi inspirado no trabalho de tipologização de FREITAS e RAMILO (2005), autores que remetem a Lopes (2001). Freitas e Ramilo trataram do uso atual do conector PORTANTO e verificaram seus empregos causais entre outros.

particularidades enunciativas das áreas estudadas, em que muitos dos conectores de causa pareciam ser utilizados como meros encadeadores ou mesmo com um sentido diferente do de causa, talvez particular daquele discurso específico.

O Gráfico 1 ilustra o número de contextos possivelmente causais e os confirmados após os testes.

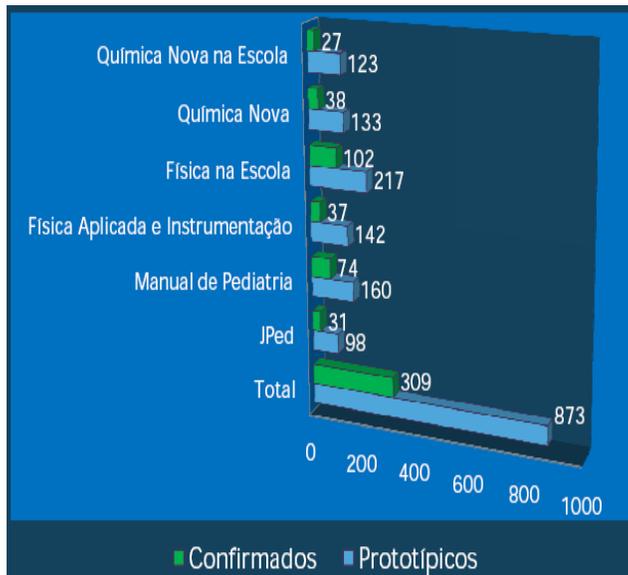



Gráfico 1: Conectores confirmados e prototípicos
(prototípicos - antes do teste; confirmados - após o teste).

Com esses dados em mãos, pudemos verificar os conectores causais mais frequentes em cada uma das três áreas contrastadas (POIS em Física e Pediatria e DEVIDO em Química), obtendo uma espécie de padronização da causalidade para cada uma delas.

Surpreendeu-nos, também, o baixo uso do conector PORQUE, um conector comum, até banal, prototipicamente causal, mas com presença quase irrelevante no *corpus* sob exame tendo-se em vista apenas um número bruto de ocorrências. Isso é indicado na Tabela 1.

Tabela 1: número bruto de ocorrências por *corpus*. O número não foi ponderado em relação à dimensão do *corpus*.

| Química | | Física | | Pediatria | |
|---|---|---|---|---|---|
| Devido | 25 | Pois | 45 | Pois | 28 |
| Assim | 18 | Devido | 23 | Assim | 20 |
| Pois | 9 | Portanto | 22 | Devido | 19 |
| Como | 3 | Como | 16 | Portanto | 18 |
| Portanto | 3 | Assim | 11 | Porque | 13 |
| Logo | 1 | Então | 9 | Como | 10 |
| Porque | 1 | Porque | 8 | Então | 0 |
| Então | 0 | Logo | 3 | Logo | 0 |

Consideramos, assim, ter encontrado um novo indício de outra particularidade da construção textual. Diferentes áreas do conhecimento apresentavam diferentes preferências no uso desses conectores. Além disso, o fato de o texto ser mais ou menos didático parecia poder influenciar uma maior presença desses conectores. Isso é que vemos nos gráficos a seguir, que mostram a frequência dos conectores estudados no *corpus* de maior didatismo e de menor didatismo. O percentual é a razão direta entre o número de ocorrências do conector e o número de *tokens* do *corpus*.

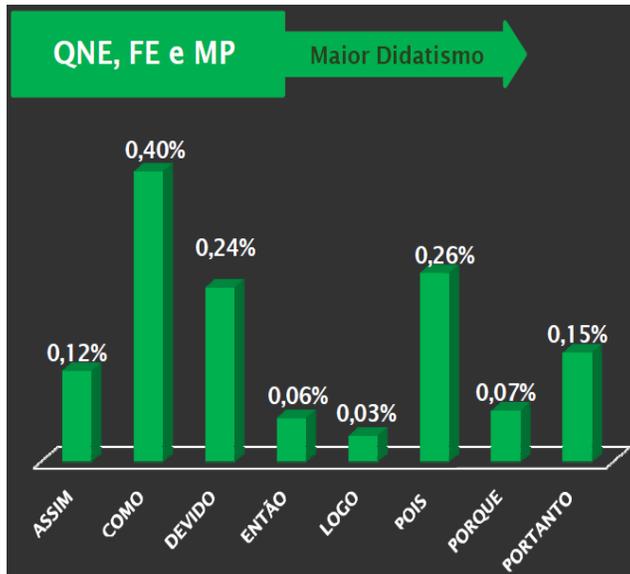



Gráfico 2: QNE = Química Nova na Escola; FE = Física na Escola; MP = Pediatria: Diagnóstico e Tratamento.

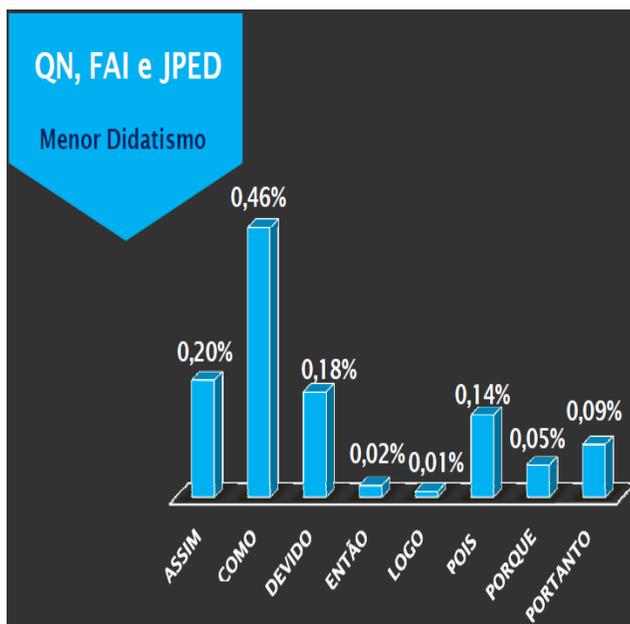



Gráfico 3: QN = Química Nova; FAI = Física Aplicada e Instrumentação; JPED = Jornal de Pediatria.

## 4. Da observação exclusiva em Pediatria

Ao aplicar os testes de topicalização na exploração anterior, enfrentamos muitas dificuldades na formulação das sentenças modificadas como também enfrentamos várias dúvidas ao compartilhar nossos julgamentos heterogêneos da causalidade nos contextos lidos. Surgiram, então, duas novas curiosidades:

a) Como os elementos cujo sentido causal foi problemático para nossa percepção seriam tratados em casos de tradução do português para o inglês?

b) Contextos de verificação menos problemática para nós teriam recebido soluções de tradução em inglês da mesma avaliação?

A partir dessas perguntas, centramos nossas observações em contextos dos conectores nos materiais bilíngues de que dispúnhamos.

### *4.1 Causalidade em tradução do português para o inglês: o caso de BECAUSE em Pediatria*

Para atender à nossa curiosidade, aproveitamos um *corpus* manualmente alinhado (português-inglês) da revista bilíngue *Jornal de Pediatria* composto por 283 artigos. O material foi generosamente cedido ao nosso grupo de pesquisa pelo tradutor R. J. Coulthard para fins de estudo. Coulthard já havia estudado esse material em seu trabalho de mestrado de 2005 (Coulthard, 2005).

Em uma primeira fase, examinamos os conectores de apenas 10 artigos desse material do *Jornal de Pediatria*. São textos originalmente escritos em português e suas respectivas versões para o inglês oferecidas pela revista ao longo do ano de 2003. Foram considerados apenas os contextos em que os conectores causais já houvessem sido submetidos a testes de causalidade. Observamos itens com confirmação de sentido causal e aqueles em que houve alguma discrepância de julgamento, mas que, ainda assim, tinham sido aprovados pelos testes efetuados. Nosso objetivo, nessa etapa, como já dissemos, foi apenas o de verificar se as escolhas tradutórias em inglês teriam mantido um sentido de causalidade mais ou menos preciso ao que percebemos nos textos originais.

Abaixo, na Ilustração 1, vemos as quantificações dos conectores e as opções tradutórias correspondentes a cada um. É importante observar que um mesmo elemento em português (língua-fonte) pode receber vários tipos de correspondentes em inglês (língua-alvo). Em verde estão assinalados os elementos de sentido causal.

Ilustração 1: conectores e opções/preferências tradutórias.

Pudemos notar que os conectores *SINCE*, *DUE TO* e *BECAUSE* foram as formas preferidas nos textos na língua-alvo. Chamaram-nos a atenção, no entanto, as escolhas que possuíam um sentido de causalidade menor ou quase nulo em relação ao da forma em português, marcado, na Ilustração 1, com cor branca. Em alguns contextos, encontramos uma elipse como tradução de um conector expresso na língua-fonte.

Enfrentando novamente dificuldade de percepção de um sentido de causalidade pelas opções presentes nos textos em inglês, visto que algumas nos pareciam desviantes do sentido de causalidade verificado na frase em português, outro questionamento surgiu:

c) Seriam essas escolhas em inglês diferentes do sentido do texto original condicionadas por alguma convenção de expressão *internacional* da área de Pediatria?

Realizamos, então, uma nova comparação. Organizamos um novo *corpus* com 10 artigos da mesma revista, mas dessa vez apenas com textos publicados em 2007. Coletamos também 10 artigos originalmente escritos em inglês publicados na revista internacional *Pediatrics: Official Journal of the American Academy* Tomamos o cuidado de obter esses 10 artigos internacionais com temáticas semelhantes as dos textos brasileiros. Isso naturalmente, rendeu-nos um bom trabalho prévio de preparação do *corpus*.

O objetivo nessa nova etapa foi verificar se as escolhas tradutórias encontradas nas versões brasileiras para o inglês correspondiam ao padrão de usos dos conectores causais em artigos de Pediatria originalmente escritos em língua inglesa.

Nessa observação, notamos que os conectores *SINCE* e *DUE TO*, escolhas recorrentes nas versões do português para o inglês, eram conectores causais de baixo uso – ou quase nulo – nos textos originalmente escritos em inglês. Já o conector *BECAUSE*, que apresentou uma frequência média baixa nas mesmas versões, mostrou-se extremamente recorrente nos contextos analisados, como podemos observar no Gráfico 4.

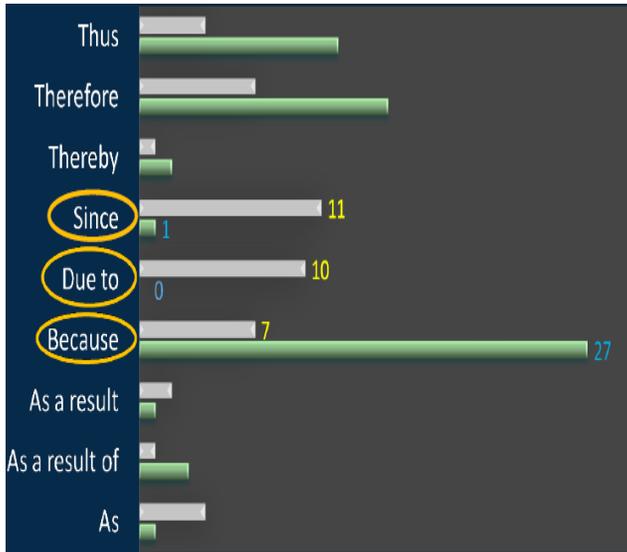



Gráfico 4: Frequências dos conectores em versões para o inglês do *Jornal de Pediatria* em artigos originalmente escritos em inglês.

Surpresas com esse resultado e querendo confirmar o que os dados coletados nas versões e textos originalmente escritos em inglês nos apresentaram, nosso *corpus* de originais em inglês foi mais uma vez ampliado, bem como o *corpus* de versões do português para o inglês.

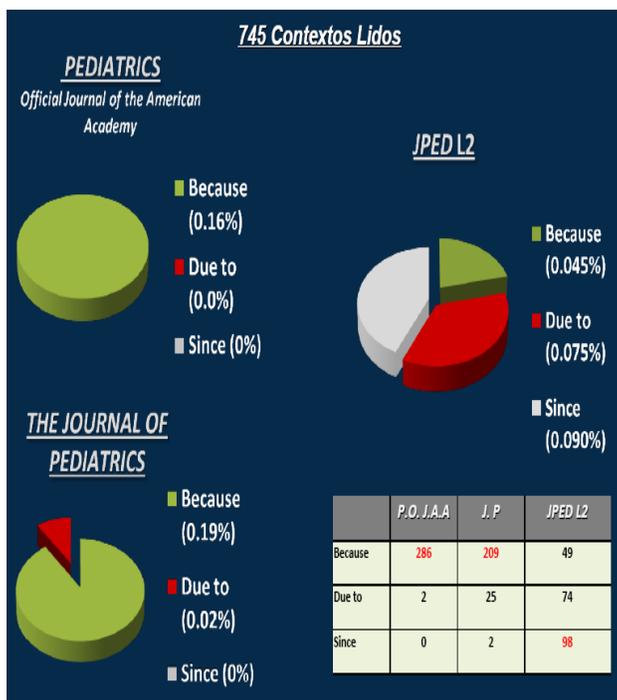
Assim, analisamos mais 37 artigos originalmente escritos em inglês de duas revistas de prestígio internacional em Pediatria: *Pediatrics: Official Journal of the American Academy* **(P. O. J. A. A)** e *The Journal of Pediatrics: Official Journal of the Association of Medical School Pediatric Department Chairs* **(J.P)**, e 37 artigos vertidos para o inglês do *Jornal de Pediatria* **(JPEDL2)**. Dessa observação, também apenas com textos de 2007, extraímos os seguintes resultados indicados a seguir na Ilustração 2 e na Tabela 2.

Illustração 2: Conectores e revistas.

Tabela 2: Conectores nos originais em inglês e nas versões do português para o inglês.

| | P.O. J.A.A | J. P | JPED L2 |
|---|---|---|---|
| Because | 286 | 209 | 49 |
| Due to | 2 | 25 | 74 |
| Since | 0 | 2 | 98 |

Nessa última etapa do trabalho de exploração, verificamos que as versões para o inglês dos artigos brasileiros ratificaram alta frequência dos conectores *SINCE* e *DUE TO*, em detrimento de *BECAUSE*.

De modo inverso, nos artigos internacionais de Pediatria originalmente escritos em inglês, *BECAUSE* foi o conector mais recorrente, enquanto *SINCE* e *DUE TO* tiveram uma frequência baixa.

Assim, pudemos inferir que o padrão internacional, aqui representado pelo *corpus* composto por artigos de Pediatria originalmente escritos em inglês, exibia indícios de preferência pelo uso do conector *BECAUSE*. Os artigos brasileiros vertidos para o inglês, por sua vez, não seguiam esse padrão, demonstrando uma convenção de expressão de causalidade bastante diferente.

Em síntese, os dados observados mostraram as seguintes tendências:

a) usos diferenciados de conectores por áreas de conhecimento;

b) *BECAUSE* como forma mais presente de conectores causais em revistas internacionais, enquanto nas revistas brasileiras há pouco uso de PORQUE;

c) a variedade de formas de conectores causais dos textos brasileiros de Pediatria em português, quando transportada para suas versões em inglês, parece gerar um padrão de inglês distinto do inglês internacional.

Essa diversidade, ao que nos pareceu, estaria assentada em convenções para expressar causa que são variadas e pouco centradas no elemento PORQUE nos artigos de Pediatria brasileiros. Essas convenções foram transferidas tal qual para a versão em inglês. Houve, naturalmente, uma formulação diferente na língua de chegada, mas as nuances de sentido e de um estilo mais rebuscado, verificadas no português pela escolha de conectores tais como, por exemplo, DEVIDO A em vez de PORQUE, foram mantidas no texto de chegada.

A leitura do tradutor da revista e as suas opções de formulação mostram que ele provavelmente não tinha acesso a esse dado sobre uma convenção preferencial para expressar causa no âmbito de revistas internacionais.

## 5. Conclusões

Com os resultados obtidos nessas sucessivas explorações, pudemos extrair indicativos importantes para dar prosseguimento aos nossos estudos, os quais podemos batizar, provisoriamente, de *Estudos de Linguagens Especializadas baseados em Corpora*. Esse

tipo de pesquisa, sucedânea da antes citada *Linguística das Linguagens Especializadas*, não despreza elementos da tessitura textual e abrange também as terminologias[10].

Dados como os obtidos, que revelaram que há padrões específicos para expressar causa em função das diferentes áreas do conhecimento e também em função dos gêneros textuais envolvidos, podem contribuir para que se reflita sobre uma nova dinâmica, também baseada em *corpora*, para qualificar o ensino de tradução em áreas especializadas como a Pediatria.

Ao examinar a expressão de causalidade em textos científicos em língua portuguesa, foi possível verificar que os usos dos conectores causais apresentam peculiaridades e particularidades em cada uma das áreas de conhecimento estudadas. Ao que parece, existem preferências das diferentes ciências por determinados conectores de causa, como também o grau de didatismo de um texto pode condicionar usos maiores ou menores de determinados elementos coesivos do tipo causal.

Quando um padrão de expressar causa entre revistas médicas brasileiras que se publicam também em inglês é diferente de um padrão de causa em inglês internacional, torna-se importante trazer esse dado para as aulas de tradução médica em favor de uma boa aceitação e do não estranhamento do texto por parte do seu público-alvo. Isso pode contribuir, conforme Tagnin (2003), para que o processo decisório da tradução gere um texto capaz de harmonizar-se com as convenções preferidas pela comunidade discursiva internacional.

Os dados obtidos nas nossas sucessivas explorações, no seu conjunto e também nas perguntas renovadas a cada etapa, são enriquecedores porque revelam uma série de inter-relações de indícios. A principal validade do que foi feito reside na mobilização gerada para novas aventuras de pesquisa. Com os devidos aprofundamentos e rigor estatístico ou mesmo que apenas de modo exploratório, restaria examinar, nos mesmos e em novos *corpora*, por exemplo, os contextos

---

[10] Para um exemplo desse tipo de trabalho dedicado à temática das terminologias em Pediatria, sugerimos consultar Silva e Feix (2008). Apresentação disponível em www.ufrgs.br/textquim, acesso via Biblioteca Virtual.

de apagamento de conectores causais no texto de chegada em inglês. Duas novas perguntas poderiam ser formuladas: Nos contextos com elipse, há ainda um sentido de causa? Havendo o sentido causal, que outros recursos teriam sido utilizados pelo tradutor para que esse sentido fosse preservado?

Trabalhos a considerar nesse tópico, por exemplo, seriam os de Gagnon e Singer (1999) sobre a detecção de inconsistências causais, além de se avaliar o rendimento das indicações de Hoffmann (1998a) sobre procedimentos estatísticos da *Linguística da Linguagens Especializadas*. A nova curiosidade seria a de saber se o sentido de causalidade, em tese desaparecido na língua de chegada, teria sido mantido por algum outro elemento ou recurso na frase e se tal apagamento seria mais comum em determinados gêneros textuais. Seria também válido observar como repercute esse tipo de procedimento de apagamento sobre a compreensão de leitura.

Para o ensino de tradução, acreditamos que estudos desse tipo são especialmente úteis para alertar sobre particularidades dos textos científicos. Particularidades que podem não ser tão evidentes, e que só serão percebidas por meio de uma observação extensiva dos quadros de usos das línguas em seus gêneros textuais.

Elementos linguísticos textuais comuns como conectores, tão menos marcados em relação aos "termos técnicos", podem ser, e usualmente são, subestimados pelos tradutores que não têm ainda uma experiência grande no domínio científico em que atuam. A consideração desses elementos de conexão frasal estará abrigada em enfoques descritivos que, tal como a LC e o enfoque estendido das Linguagens Especializadas aqui apresentado, considerem *a priori* tudo importante nos textos-*corpora*.

A extensa vivência do tradutor experiente ou sua condição de *expertise* equivaleria a já ter podido conhecer grandes *corpora*, acumulando um conhecimento igualmente extensivo sobre as convenções das linguagens e das culturas de escrita com que lida. Mas, sem um acesso prévio e real, direto ou indireto, a esses *corpora* especialmente reunidos, o tradutor somente teria tal "sabedoria" a partir de um ou mais textos, conhecidos aos poucos e a cada dia, ao longo dos vários anos de sua prática diária.

Para além de um aproveitamento mais imediato do que buscamos exemplificar com as explorações relatadas, elas têm, sem dúvida, uma boa capacidade para provocar inúmeras novas curiosidades de pesquisa. Assim, como já afirmou Santos (2007), embora os *corpora* não permitam responder a todos os problemas, são uma fonte de ideias e de hipóteses incansável e abundante.

A reunião da nossa abordagem mais ampla sobre linguagens científicas, a que denominamos aqui *Estudos de Linguagens Especializadas baseados em Corpora,* com a LC produz também a crença prévia e fundamental de que as respostas para quaisquer perguntas colocadas sobre a linguagem científicas em uso podem estar nos próprios *corpora*. Encontrar as respostas e relacionar a elas novas indagações dependerá da nossa capacidade de escolher os melhores acervos de dados e de saber navegar em meio aos tantos mares de língua[11] que os *corpora* nos oferecem.

## Agradecimentos

Este trabalho foi desenvolvido com apoio do **CNPq**, Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil.

## Referências

### Dos *corpora*

FERREIRA, José Paulo (Org.). 2005. Pediatria: diagnóstico e tratamento. Artmed.

JORNAL DE PEDIATRIA (JPed). 2003. Disponível em: www.jped.com.br. Acesso em julho de 2008.

PEDIATRICS: Official Journal of the American Academy of Pediatrics. 2007/2008. Disponível em: http://pediatrics.aappublications.org/. Acesso em julho 2008.

REVISTA FÍSICA NA ESCOLA. 2007. SBF. Disponível em: www.sbfisica.org.br. Acesso em junho de 2008.

---

[11] Plagiamos aqui uma bela figura de linguagem de autoria de Diana Santos (Santos, 2007).

REVISTA DE FÍSICA APLICADA E INSTRUMENTAÇÃO. 2003/2005. Disponível em: www.sbfisica.org.br. Acesso em julho de 2008.

REVISTA QUÍMICA NOVA. 2004. FAPESP. Disponível em: www.quimicanova.sbq.org.br. Acesso em julho de 2008.

REVISTA QUÍMICA NOVA NA ESCOLA. 2004. FAPESP. Disponível em: www.quimicanova.sbq.org.br. Acesso em julho de 2008.

THE JOURNAL OF PEDIATRICS: Official Journal of the Association of Medical School pediatric department chairs. 2007/2008. Disponível em: www.amspdc.org. Acesso em julho de 2008.

## Das obras e artigos citados

ALLE, Cybele M. O. ; Evers, Aline. 2008 Leituras da causalidade na tradução em Pediatria: elementos de conexão frasal. XX Salão de Iniciação Científica. UFRGS.

ALMEIDA, G.M.B.; Aluísio, S.M.; Teline, M.F. Extração manual e automática de terminologia: comparando abordagens e critérios . In: 1o. Workshop em Tecnologia da Informação e da Linguagem Humana, 2003, São Carlos. Anais do TIL'2003.

CABRÉ, María Teresa. 2001. La Terminologia. Representación y comunicación. IULA. ISBN: 978-84-9674259-8.

CONDAMINES, Anne. 2005. Linguistique de corpus et terminologie. In: Langages, 157, p. 36-47. Disponível em: http://w3.univ-tlse2.fr:8880/erss/index.jsp?perso=acondami&subURL=Langagesdef.pdf. Acesso em junho de 2008.

EVERS, Aline ; Alle, C. M. O. ; Marcolin, Paula . Causalidade expressa via conectores em Química, Física e Pediatria: um estudo exploratório. In: XX Salão de Iniciação Científica, 2008. UFRGS.

FINATTO, Maria José Bocorny. 2004. Termos, textos e textos com termos: novos enfoques dos estudos terminológicos de perspectiva linguística. In: Isquerdo, Aparecida Negri; Maria da Graça Krieger (Orgs.). As ciências do léxico: lexicologia, lexicografia, terminologia, volume II. Ed. UFMS. ISBN: 85-7613-034-3.

FINATTO, Maria José Bocorny; José Cláudio Del Pino; Rafaela Guimarães Barbosa e Estela Rubia Brugalli Corbellini. 2006. O vocabulário além da Terminologia: expressões de causalidade em textos de Química. In: MATRAGA. Revista do Programa de Pós-

Graduação em Letras/ UERJ. Ano 13, nº19. Editora Caetés. ISBN: 1414-7165.

FINATTO, Maria José Bocorny; Siane Simioni. 2007. A causalidade no texto de Química: efeitos da tradução e estrutura coesiva. In: Calidoscópio: Revista de Linguística Aplicada vol. 05. Unisinos. ISBN: 16798740.

FINATTO, Maria José Bocorny. 2007(a). Exploração terminológica com apoio informatizado: perspectivas, desafios e limites. In: Isquerdo, Aparecida Negri; Maria da Graça Krieger (Orgs.). As ciências do léxico: lexicologia, lexicografia, terminologia, volume III. Ed. UFMS. ISBN: 978-85-7613-102-1.

FINATTO, Maria José Bocorny. 2007(b). Exploração terminológica com apoio informatizado: diálogos entre terminologia e linguística de corpus. In: Mercè Lorente; Rosa Estopà; Judit Freixa; Jaume Martí ; Carles Tebé. (Org.). Estudis de linguística i de linguística aplicada en honor de M. Teresa Cabré Castellví. Institut Universitari de Linguística Aplicada de la Universitat Pompeu Fabra, v. 2, p. 221-230. ISBN: 978-84-967.

GAGNON, Nathalie e Murray Singer. 1999. Detecting causal inconsistencies in scientific text. In: Susan R. Goldman; Arthur C. Graesser e Paul van den Broek (Eds.). Narrative Comprehension, Causality and Coherence: Essays in Honor of Tom Trabasso. Mahwah. Lawrence Erlbaum Associates. p. 179. ISBN: 0-8058-3358-7.

GAUDIN, François. 1993. Pour une socioterminologie. Des problèmess semantiques aux pratiques institutionnelles. Publications de l'Université de Rouen. ISBN: 2–8011–1319–0.

HOFFMANN, Lothar. 1988a. Vom Fachwort zum Fachtext: Beiträge zur Angewandten Linguistik. Gunter Narr Verlag. ISBN: 978-3-87808-699-4.

HOFFMANN, Lothar. 1988b. Grundbegriffe der Fachsprachenlinguistik. In: Germanistisches Jahrbuch für Nordeuropa. 7. Folge. Deutsche Fachsprachen in Forschung und Lehre. Helsinki, Estocolmo.

HOFFMANN, Lothar. 1998. Syntaktische und morphologische Eigenschaften von Fachsprachen. In: Lothar Hoffmann et al. (Orgs.) Fachsprachen: ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft. Walter de Gruyter.

HOFFMANN, Lothar. 2004. Fachsprache / Language for Special Purposes. In: Ammon, Ulrich et al. (Orgs.) Sociolinguistics: an international handbook of the science of language and society /

Soziolinguistik: ein internationals Handbuch zur Wissenschaft von Sprache und Gesellschaft. Walter de Gruyter, vol. 1, 2ª ed. revisada e ampliada, edição especial. ISBN-13: 978-3110158847.

HOFFMANN, Lothar. 2004. Conceitos básicos da linguística das linguagens especializadas. Traduzido por Maria José Bocorny Finatto. In: Cadernos de Tradução, n.17 – out./dez., UFRGS, p. 79-90. ISBN: 03597210.

KALVERKÄMPER, Hartwig. 1983. Textuelle Fachsprachen-Linguistik als Aufgabe. In: Zeitschrift für Literaturwissenschaft und Linguistik, v. 51/52, n° 13, p. 124-166. ISBN: 0049-8653.

KOCOUREK, Rostilav. 1991. Textes et Termes. In: Meta, vol. 36, n.1, mars, Numéro Spécial. La Terminologie dans le monde: orientations et recherches, p.71-76.

KRIEGER, Maria da Graça. 2008. Por que Lexicografia e Terminologia: relações textuais? In: Finger, Ingrid & Gisela Collischonn (Orgs.). Anais do 8° Encontro do Círculo de Estudos Lingüísticos do Sul - CELSUL, Porto Alegre, de 29 a 31 de outubro de 2008. Pelotas: EDUCAT. ISBN: 978-85-7590-115-1.

LEITCHIK, Vladimir M. 2002. Elements Of Terminological Text Theory. In: Shelov, S.D.; Leitchik, V.M. (eds.) Russian Terminology Science (1992-2002). (col. of H. Picht e C. Galinski). Termnet Publisher. ISBN 3-89913-409-5.

LOPES, Ana Cristina Macário et al. 2001. As construções com portanto no PE e no PB. In: Scripta, volume 5, número 9. PUC Minas.

NEVES, Maria Helena de M.. 2000. Gramática de usos do Português. Editora UNESP. ISBN: 8571392889)

NEVES, Maria Helena de M.. 1999. Gramática do Português falado. Volume VI: Novos estudos. Humanitas FFLCH/USP.

NEVES, Maria Helena de M.; Souza, E.M.; Domingos, E. As construções causais. In: Neves, Maria H. M. Gramática do português falado. Volume II: Novos estudos. São Paulo: Humanitas FFLCH/USP, 1999.p. 49-497. ISBN:85-8607-45-09.

RAJAGOPALAN, Kanavillil. 2007. Panorama da Linguística de Corpus (LC) In : Gerber, Regina Márcia; Vera Vasilévski (Orgs.). Um percurso para pesquisas em base em corpus. Editora da UFSC. ISBN: 978-8532-80-409-9.

ROLIM, Aline Marques. 2000. A reciclagem de resíduos plásticos pós-consumo em oito empresas do Rio Grande do Sul. Dissertação (Mestrado em Administração) - Universidade Federal do Rio Grande do Sul, Orientador: Luis Felipe Machado do Nascimento.

SAGER, Juan Carlos. 2007. La evolución de los lenguajes de especialidad y la terminología. In: Estudis de linguística i de linguística aplicada en honor de María Teresa Cabré Castellví. Vol. I - De mestres. IULA/UPF, p. 105-122. ISBN: 978-84-96742-08-6.

SANTOS, Diana. 2007. Breves explorações num mar de língua, Ilha do Desterro **52**, 1. UFSC. ISBN: 0101-4846.

Sardinha, Tony Berber. 2004. Linguística de Corpus. Manole. ISBN: 852-0416-764.

SILVA, Maria Cristina Alencar; Gustavo Azambuja Feix. 2008. Reconhecimento inicial de usos do termo prevalência em Pediatria: subsídio para um catálogo de expressões recorrentes na área. XX Salão de Iniciação Científica. UFRGS.

Sinclair, John M.. 1991. Corpus, Concordance, Collocation. Oxford University Press. ISBN: 978-0194-3-7144-5.

STUBBS, Michael. 1996. Text and corpus analysis: computer-assisted studies of language and culture. Blackwell. 227p. (Language & Society series, v.23). ISBN: 978-0-631-19512-2.

STUBBS, Michael.. 2001. Words and phrases: Corpus studies of lexical semantics. Blackwell. ISBN: 978-0-631-20832-7.

SWALES, John M. 1990. Genre analysis: English in academic and research settings. Cambridge University Press. ISBN: 978-05-2133-813-4.

TEMMERMANN, Rita. 2000. Towards New Ways of Terminology Description. The sociocognitive-approach. John Benjamins Pub. Co. ISBN: 978-15-5619-772-7.

# PROCESSAMENTO DE LINGUAGEM NATURAL E O TRATAMENTO COMPUTACIONAL DE LINGUAGENS CIENTÍFICAS.

*Renata Vieira*

*Lucelene Lopes*

*FACIN-PUCRS*

**Resumo:** *Desde o surgimento das técnicas de Processamento de Linguagem Natural - PLN, muitos avanços foram obtidos, mas a compreensão plena de linguagem natural por métodos computacionais está ainda longe de ser resolvida. Apesar disso, o tratamento computacional da língua e bases textuais é um campo muito promissor. Em particular, os usuários das linguagens específicas de domínio, como as linguagens científicas, podem se beneficiar de PLN, como por exemplo, na busca por informação em coleções de documentos. Este capítulo apresenta a área de PLN, fornecendo um breve histórico, objetivos, métodos e técnicas computacionais e as suas possíveis aplicações, que incluem, por exemplo, a tradução automática, a categorização de textos, o processamento de discurso e a web semântica, que pode ser considerada como o foco atual da área. Com o desafio da web semântica surge a necessidade de representar formalmente o conhecimento de domínios específicos através de ontologias. Dessa forma, apresenta-se também uma definição sobre ontologias e exemplos para facilitar a compreensão do leitor. Percebe-se através da leitura, que a construção automática de ontologias a partir do conhecimento representado em textos, não é uma tarefa simples. Este processo, denominado Aprendizagem de Ontologias, é composto de cinco etapas, sendo que a primeira, denominada Extração de Termos, é abordada em maior detalhe neste capítulo. Nesse contexto, são apresentadas abordagens linguística, estatística e híbrida para a extração de termos de bases textuais (corpus). Através de trabalhos relacionados, percebe-se que a extração de termos a partir de corpora em linguagem científica é bastante relevante. Logo, o tratamento*

*computacional de textos científicos torna-se viável por tratar-se de uma linguagem objetiva que segue padrões linguísticos com a finalidade da circulação do saber. Evidentemente, esse é um tópico de pesquisa em aberto e particularmente grande são as expectativas quando se trata de trabalhos sobre a língua portuguesa. Por esta razão, este capítulo não esgota o tema. No entanto, o panorama histórico apresentado e a contextualização dos problemas a serem atacados servem para delinear o campo onde os avanços científicos e tecnológicos devem ocorrer nos próximos anos. Desta forma, espera-se com este capítulo motivar o leitor a contribuir nos esforços de pesquisa de PLN nas áreas de manipulação de textos científicos, extração de termos a partir de corpora e, consequentemente, aprendizagem de ontologias.*

## Introdução

Processamento de Linguagem Natural (PLN) é uma área de Ciência da Computação que estuda o desenvolvimento de programas de computador que analisam, reconhecem e/ou geram textos em linguagens humanas, ou linguagens naturais.

O PLN não é uma tarefa trivial devido à rica ambiguidade da linguagem natural. Essa ambiguidade torna o PLN diferente do processamento das linguagens de programação de computador, as quais são formalmente definidas evitando, justamente, a ambiguidade.

Segundo (Liddy, 2003), alguns objetivos usuais em PLN são (a) recuperação de informação a partir de textos, (b) tradução automática, (c) interpretação de textos e (d) realização de inferências a partir de textos. No entanto, diversos esforços atuais nesta área congregam vários destes objetivos, como é, por exemplo, o caso de construção de ontologias a partir de textos (Cimiano, 2006). Em (Lopes et alli, 2009a) o trabalho reporta experimentos nessa área, onde é necessário recuperar informação (a) e fazer inferências (b). Outro exemplo é (Papineni et alli, 2002), onde técnicas de interpretação de texto (a) e recuperação de informações (b) para qualificar o processo de tradução automática são utilizadas.

Vários métodos ou técnicas computacionais podem ser usados para analisar e interpretar a linguagem natural. Existe uma clara

distinção, por exemplo, entre as técnicas empregadas na linguagem falada e na linguagem escrita, mas também além destas divisões é possível ver diferenças claras entre as técnicas necessárias para diferentes tipos de aplicação, e que podem envolver diferentes tipos de textos, como por exemplo, textos científicos, jornalísticos, literários, etc. Neste sentido, a grande maioria das aplicações é voltada para um tipo específico de linguagem. Um dos sistemas pioneiros da área, o Eliza (Weizenbaum, 1966), simula o diálogo do usuário como se ele fosse o paciente de um psicólogo virtual. Neste programa, que data de 1966, as técnicas de PLN são bastante simplistas, mas servem para ilustrar uma das primeiras aplicações que, ainda hoje, é bastante popular nas aplicações do tipo *chatbot*.

De um ponto de vista linguístico, o foco das pesquisas em PLN pode estar em um de cinco níveis de análise: (a) fonético ou fonológico, (b) morfológico, (c) sintático, (d) semântico ou (e) pragmático. Todos esses níveis possuem suas características próprias e suas dificuldades associadas, mas cada aplicação de PLN pode ter a preocupação mais voltada para um subconjunto desses níveis. Por exemplo, aplicações sobre textos científicos usualmente não têm preocupação com uma análise fonológica (a), por outro lado, aplicações que façam uma interface com reconhecimento de voz focam esse nível de análise.

Associado ao tratamento de linguagens científicas pode-se destacar a relevância das atividades de reconhecimento de entidades nomeadas, identificação de termos, extração de informação, entre outras. Todas as atividades complexas podem combinar análise morfossintática e semântica, que podem ou não estar associadas a recursos linguísticos externos, como tesauros por exemplo. De um ponto de vista prático, muitas aplicações de PLN escolhem se dedicar a apenas alguns níveis para reduzir a complexidade de tratamento.

## 1. Histórico da Área de Processamento de Linguagem Natural – PLN

Historicamente, a área de PLN começou com tentativas de tradução automática na segunda metade da década de 1940 (Weaver, 1949). Esses trabalhos iniciais estavam relacionados com esforços

prévios de quebra de códigos durante a Segunda Guerra Mundial. De um ponto de vista teórico, esses trabalhos inicias em tradução automática estavam baseados na criptografia e teoria da informação (Shannon, 1949). Em 1957 Chomsky desenvolveu trabalhos relevantes sobre o tema. Um trabalho particularmente relevante desta época é o livro *Syntactic Structures* (Chomsky, 1957) que introduziu a gramática gerativa, a partir daí ficou mais clara como a área de linguística poderia auxiliar a área de tradução automática.

Nessa época houve também a inclusão de outras aplicações de PLN, especialmente a do reconhecimento da fala (*speech recognition*). Nesta área houve a primeira grande divergência que de certa forma permanece até hoje, pois parte da comunidade optou pelo uso de linguística teórica e parte optou por métodos estatísticos. Infelizmente, cada uma destas partes rechaçava os métodos da outra parte prejudicando a integração destas duas abordagens.

Esse período marca também o advento da Teoria Sintática da Linguagem (Chomsky, 1965) e dos Algoritmos de Parsing (Aho e Ullman, 1972). Estes avanços foram muito importantes para a área, ainda que na época estes avanços tenham sido recebidos com um entusiasmo excessivo de que em poucos anos tradutores automáticos perfeitos estariam disponíveis. Esta expectativa se mostrou indevida tanto pelos conhecimentos linguísticos e computacionais da época, quanto por uma impossibilidade teórica da tarefa de tradução automática perfeita (Bar-Hillel, 1960).

Na verdade, consequência disto ou não, em 1966 o comitê assessor para processamento automático da língua (ALPAC) da academia americana de ciência recomendou que esta área não recebesse mais financiamento governamental, pois a tradução automática estava muito aquém dos conhecimentos científicos atuais. Em contraste com esta decisão, vários avanços teóricos e práticos foram feitos nos anos seguintes. Entre eles pode ser citado o trabalho teórico de Chomsky que introduziu o modelo computacional de competência linguística (Chomsky, 1965) que resultou nas gramáticas gerativas transformacionais. Diversos trabalhos subsequentes tentaram aproximar estes conceitos de modelos computacionalmente tratáveis.

A partir desse período, houve uma multiplicação dos estudos sobre PLN com o estabelecimento de diversas subáreas que vem sendo pesquisadas até hoje. Estas áreas se dedicam a assuntos tão variados quanto à categorização de textos e à extração de informações, passando pelos tradicionais temas de tradução automática e sistemas de diálogo. Os trabalhos desenvolvidos neste período podem, segundo diversos autores (Jurafsky e Martin, 2009), ser agrupados em quatro grupos segundo os paradigmas utilizados: os métodos estocásticos, os métodos baseados em lógica, os métodos de entendimento de linguagem natural, e os métodos de modelagem de discurso.

Os trabalhos do grupo de métodos estocásticos são baseados em abordagens estatísticas e frequentemente utilizam formalismos com os modelos ocultos de Markov (*HMM – Hidden Markov Models*). Estes métodos estão na base de diversos trabalhos de reconhecimento e síntese de fala (Rabiner e Juang, 1986). Estes trabalhos estão na origem dos atuais trabalhos em que métodos estatísticos são empregados para diversas aplicações de PLN.

Os trabalhos baseados em lógica começaram com *Q-systems* e gramáticas metamórficas (Colmeraeur, 1970 e 1975) que foram os precursores da linguagem Prolog (Colmerauer, 1996) e das gramáticas denitivas de cláusula (Pereira e Warren, 1980). Desta mesma época datam também as iniciativas de gramáticas funcionais na sua versão inicial (Kay e McDaniel 1979) e na versão léxica (Bresnan e Kaplan, 1982) que deixaram clara a importância da unificação da estrutura das características.

Os trabalhos do grupo baseado em entendimento da linguagem natural continuaram na vertente de entendimento do discurso. De um ponto de vista teórico, são típicos deste grupo os trabalhos sobre Gramáticas de Caso (Fillmore, 1968), Redes Semânticas (Quillian, 1968), Teoria de Dependência Conceitual (Schank, 1972), Redes de Transição Aumentada (Woods, 1970) e Semântica de Preferência (Wilks, 1975). De um ponto de vista prático, este período viu o aparecimento de diversos programas que faziam uso de PLN. Este foi o caso dos sistemas de diálogo ELIZA (Weizenbaum, 1966) e PARRY (Colby, 1973), mas também os sistemas de reconhecimento de fala

SHRDLU (Winograd, 1971), LUNAR (Woods, 1977), LIFER/LADDER (Hendrix, 1977) e PLANES (Waltz, 1978).

Finalmente, o grupo focado na modelagem do discurso voltou suas atenções a questões semânticas. Trabalhos significativos desta época como (Grosz, 1977) visavam diálogos funcionais (diálogos que especificam uma tarefa a ser executada). Os trabalhos subsequentes de Gennari em particular definem uma teoria de partição do discurso baseado em relações entre a estrutura da tarefa a executar e a estrutura do diálogo que descreve esta tarefa (Gennari et alli, 1986). Nesta mesma época foi desenvolvida por Mann e Thompson uma Teoria de Estrutura Retórica (Mann e Thompson, 1988) que atribui uma estrutura hierárquica para o discurso com o intuito de geração automática de texto. Outros trabalhos deste período também foram dedicados à geração de linguagem natural, como é o caso dos geradores de resposta TEXT (McKeown, 1985) e MUMMBLE (Mcdonald, 1983) que usam predicados retóricos para produzir descrições declarativas na forma de parágrafos.

Desde o início da década de 1990, o crescimento da internet e a profusão de textos disponíveis direcionaram os esforços do PLN para o tratamento de textos mais do que para o discurso falado. Nesta época iniciaram as pesquisas sobre *corpora* anotados sintaticamente, ou seja, conjuntos de textos sobre um domínio de conhecimento onde cada uma das suas palavras foram identificadas segundo sua função sintática. Vários destes trabalhos foram desenvolvidos para a língua inglesa, sobre três *corpora* bastante populares: *Brown corpus* (Kucera e Francis, 1967), *Lancaster-Olso-Bergen corpus* (Garside et alli, 1987) e *Penn Treebank* (Marcus et alli, 1993). Neste sentido, abordagens simbólicas baseadas em conceitos linguísticos utilizadas em conjunto com abordagens estatísticas possibilitaram resultados mais robustos (Manning e Shütze, 1999). Na verdade, esta reconciliação entre métodos linguísticos e estatísticos que se percebe atualmente desfaz a divisão de abordagens feita na área quase 50 anos atrás.

A grande quantidade de informação a ser tratada que impulsionou a reconciliação dos métodos estatísticos e linguísticos teve desde a virada do século uma outra consequência interessante com a incorporação de técnicas de aprendizado de máquina (Witten e Frank,

2005). As técnicas de aprendizado de máquina são particularmente aplicáveis no contexto de conjuntos de dados humanamente intratáveis, mas de onde se pode inferir padrões e, consequentemente, informação. Naturalmente, os últimos anos têm testemunhado uma convergência das técnicas de PLN baseadas em *corpus* com técnicas de aprendizagem de máquina, ou mais especificamente, técnicas de mineração de dados.

Este aumento significativo das ferramentas à disposição dos pesquisadores de PLN permitiu também um aumento significativo nas ambições da área. Retomou-se seriamente os trabalhos de tradução automática, apesar da consciência de ser inatingível uma tradução perfeita. Uma quantidade muito grande de tradutores automáticos está disponível na internet como é o caso dos sites especializados como o *babelfish* (http://babelfish.yahoo.com), mas também de serviços de tradução embutidos como os disponíveis automaticamente pelo gigante de pesquisas *Google* (http://www.google.com/language_tools).

Igualmente, a evolução de outras áreas da computação, como é o caso da computação pervasiva (recursos computacionais presentes em atividades cotidianas) ou aplicações web, abriu espaço para novos sistemas de reconhecimento e geração automática da linguagem. Alguns sistemas recentes utilizam PLN com o objetivo de responder perguntas de forma clara e direta através de conhecimento semântico. Exemplos práticos deste tipo de aplicação, disponíveis na internet, são os sistemas *Ask* (http://www.ask.com), *Lexxe* (http://www.lexxe.com), *Powerset* (http://www.powerset.com) e *Hakia* (http://www.hakia.com). Outro exemplo desenvolvido com o mesmo princípio é o sistema *True Knowledge* (http://www.trueknowledge.com), que ao contrário dos demais que direciona a vários links relacionados, o *True Knowledge* com perguntas simples feitas pelo usuário, insere a resposta direta, além de também fornecer *links* relacionados. Esse sistema permite que os usuários acrescentem informações, tornando-o cada vez mais completo e acurado.

Porém, talvez o objetivo mais ambicioso do processamento de linguagem natural reside atualmente na construção da web semântica que pretende estabelecer uma ponte entre o enorme volume de dados disponível na internet e as demandas de seus milhões de usuários

(Berners-Lee et alli, 2001). A web semântica é uma iniciativa que busca identificar e representar o significado de páginas na web de forma que tanto pessoas como máquinas possam identificá-los. Neste sentido, o grande desafio é a representação do conhecimento em um formato adequado, que neste contexto é feito através de Ontologias (Gruber, 1995).

## 2. Ontologias para Domínios Específicos

Ontologias são representações formais de um modelo de domínio. Geralmente, uma ontologia é entendida como um conjunto de conceitos organizados hierarquicamente, um conjunto de relações e um conjunto de atributos. A definição formal de ontologias foge ao escopo deste trabalho, mas o leitor interessado pode encontrar um extenso material em (Ehrig, 2007) e (Buitelaar et alli, 2005). O importante aqui é ter em mente o tipo de representação que é feito por uma ontologia. Para facilitar o entendimento, vamos considerar uma ontologia que represente um conjunto de pessoas e suas relações familiares. Para esta ontologia exemplo, vamos admitir os seguintes conceitos: *pessoas*, *homens*, *mulheres*, *meninos* e *meninas* organizados hierarquicamente conforme a Figura 1. Neste exemplo, todos homens e mulheres são pessoas, todos meninos são homens e todas meninas são mulheres. Esta hierarquia de conceitos é representada na Figura 1 por arcos com linhas contínuas.

Para esta ontologia vamos considerar um conjunto de relações *pai_de*, *mãe_de*, *filho_de* e *filha_de*. Estas relações estão indicadas na Figura 1 por arcos pontilhados. As relações *pai_de* e *filho_de* existem somente entre um homem e qualquer pessoa e analogamente só uma mulher pode ser *mãe_de* e *filha_de* uma pessoa. Completando o núcleo desta ontologia, podemos associar atributos aos conceitos desta ontologia. Por exemplo, podemos associar a cada pessoa um atributo numérico que indique sua idade. Este atributo é comum a todas as pessoas, logo na Figura 1 este atributo, *idade*, está representado associado ao conceito *pessoas*.

De acordo com a aplicação pretendida, e também com os autores consultados, a definição de uma ontologia pode se limitar a

descrever a sua estrutura (conceitos, relações e atributos e suas respectivas hierarquias) como descrito na Figura 1. Em outros casos, a ontologia só é considerada completa quando, além da estrutura, também é necessário popular os conceitos com instâncias. Um exemplo destes casos, para o exemplo da Figura 1, seria associar diversas pessoas a cada um dos conceitos. Este processo de associar instâncias a uma estrutura de uma ontologia é chamado de população de ontologias (Cimiano, 2006).

Uma vez um domínio representado em uma ontologia, é possível identificar semanticamente as consultas feitas por usuários, tanto quanto classificar as páginas existentes segundo seus significados. Na verdade, diversos recursos podem ser semanticamente identificados. Porém o processo de construção de uma ontologia é, em geral, lento e bastante complexo. E por isso existem diversas técnicas e ferramentas para a construção de ontologias. No entanto, dada a complexidade do processo, é difícil imaginar a construção de uma ontologia relevante sem o auxílio de ferramentas computacionais. Curiosamente, pela mesma razão ainda não existem sistemas capazes de construir uma ontologia de forma completamente automática.

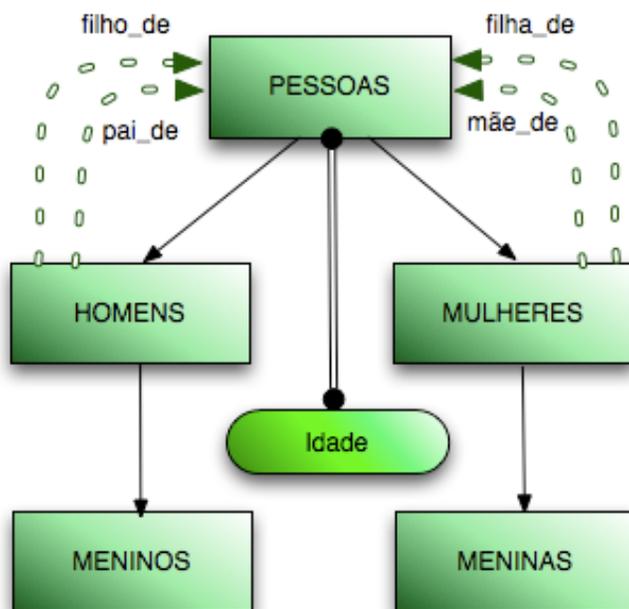



Figura 1 Exemplo de Estrutura de uma Ontologia

Os esforços semiautomáticos mais simples são baseados na utilização de uma ferramenta que permita organizar ontologias que devem ser projetadas por um usuário que conheça o domínio a ser descrito pela ontologia. Dentre estas ferramentas, provavelmente a mais popular é o *Protégé* (Crubézyb et alli, 2003), uma ferramenta que permite ao usuário construir e manipular ontologias quase que só manualmente, pois apenas algumas verificações e visualizações automáticas são possíveis. Porém, uma das funcionalidades muito interessantes do *Protégé* é a possibilidade de adicionar *plugins* capazes de realizar operações sobre ontologias. Outra funcionalidade interessante do *Protégé* é a possibilidade de modelar ontologias segundo dois protocolos: OKBC - *Open Knowledge Base Connectivity* (OKBC, 2009) e OWL *Web Ontology Language* (Mcguinness e Van Harmelen, 2009). Maiores detalhes sobre a ferramenta *Protégé* podem ser encontrados em http://protege.stanford.edu.

Outra ferramenta semiautomática de construção de ontologias é o OntoGen (Fortuna, 2007) que combina técnicas de mineração de textos com uma interface de utilização que facilita a escolha dos conceitos e relações. Na verdade, o OntoGen, parte de um *corpus* e oferece ao usuário conjuntos de termos candidatos a conceitos e cabe ao usuário estabelecer a hierarquia entre os conceitos, bem como as relações entre eles. Neste sentido, esta ferramenta é mais uma ferramenta de edição de ontologias que possui um processo de extração de termos candidatos integrado. Maiores detalhes da ferramenta OntoGen podem ser encontrados em http://ontogen.ijs.si.

O primeiro problema para gerar ontologias a partir de textos é identificar quais tarefas são necessárias para a construção efetiva de uma ontologia. Ainda que de acordo com os objetivos, os componentes de uma ontologia possam variar, todas as possibilidades sempre consideram pelo menos um conjunto de conceitos e sua hierarquia (uma taxonomia) e relações entre estes conceitos.

Para adquirir esse conhecimento conceitual dos textos (*corpus*), aplicam-se vários métodos e técnicas da área de Inteligência Artificial que ajudam a automatizar o processo de construção de ontologias. Este processo de construção automática de ontologias a partir de textos denomina-se Aprendizagem de Ontologias (Buitelaar et alli, 2005). Segundo Buitelaar, este processo divide-se em cinco etapas básicas: extração de termos candidatos a conceitos de um domínio; determinação de sinônimos entre os termos candidatos e escolha dos conceitos; identificação da relação hierárquica entre os conceitos; identificação de relações entre os conceitos; e identificação de instâncias (população da ontologia). A aprendizagem de ontologias pode ser representada em camadas de acordo com a Figura 2.

Figura 2 Etapas de Aprendizagem de Ontologias

Logicamente, os passos descritos na Figura 2 devem ser executados sequencialmente, sendo a extração de termos candidatos a conceitos a primeira e mais importante tarefa, pois da qualidade dos resultados desta etapa depende a qualidade dos resultados de todas as demais etapas. Note-se que esta afirmação não significa que as outras etapas sejam mais simples, ou que não seja necessária uma grande preocupação com a eficiência de cada uma delas. A qualidade da ontologia é completamente dependente das demais tarefas, porém, caso a extração de termos candidatos seja deficiente, o resultado de todas as demais etapas não poderá compensar esta deficiência. Por esta razão, neste capítulo focamos nessa etapa do processo de aprendizagem de ontologias.

## 3. Extração de Termos Candidatos a Conceitos

A extração de termos pode ser feita segundo três abordagens: linguística, estatística e híbrida.

A **abordagem linguística** faz uso de um nível avançado de processamento linguístico, que geralmente envolve a identificação da organização das frases dos textos (*corpus*) sintaticamente. Para essa análise utilizam-se programas de computador denominados *parsers* que fazem anotação sintática, e ocasionalmente também semântica, de maneira automática. A partir da anotação feita pelo *parser*, outros programas utilizam essas anotações (geralmente descritas em arquivos XML) como input adicional para a tarefa de identificação dos termos candidatos a conceitos de uma ontologia. Uma informação linguística relevante para extrair automaticamente os termos de um *corpus* é a identificação dos sintagmas nominais, isso permite a extração apenas de unidades de sentido previamente identificadas pelo *parser*. Um sintagma nominal é um conjunto de palavras que exerce a mesma função gramatical que um substantivo. Segundo (Kuramoto, 2002), ao contrário das palavras isoladas cujo significado depende fortemente do contexto, sintagmas nominais tendem a permanecer com o mesmo significado.

Na **abordagem estatística** os candidatos a termos são extraídos segundo uma análise da sua frequência de ocorrência no *corpus*. Existem ferramentas específicas que utilizam essa abordagem, por exemplo, o NSP – *Ngrams Statistic Package* (Banberjee, 2003), que possui um conjunto de programas que realizam a tarefa de extração de termos de um *corpus* de maneira totalmente estatística. Esse método de *Ngrams* extrai termos composto por um número fixo de palavras. Usualmente buscam-se bigramas (termos compostos por duas palavras) e trigramas (termos compostos por três palavras), porém termos com qualquer número de palavras, inclusive unigramas, podem ser extraído.

A **abordagem híbrida** é a uma combinação das duas outras abordagens, onde tanto conceitos linguísticos, quanto informações estatísticas são consideradas. Na verdade, apesar das diferenças históricas entre as abordagens linguísticas e estatísticas citadas

anteriormente, a grande maioria dos métodos de extração de termos pode sempre ser visto como um método híbrido.

A extração de termos de *corpus* em linguagens científicas é um tema bastante relevante abordado em vários trabalhos de pesquisa como é o caso do trabalho de Lopes et alli, 2009b, que faz uma comparação entre duas abordagens, a linguística e a estatística. Neste trabalho, são extraídos sintagmas nominais utilizando uma abordagem linguística e o resultado é comparado com a extração utilizando uma abordagem eminentemente estatística. Em ambos os casos são utilizadas ferramentas automáticas de extração de termos: NSP (Banberjee, 2003) para a abordagem estatística e EXATOLP (Lopes et alli, 2009c) para a abordagem linguística. Nesta comparação todo o processo de extração é realizado a partir de um *corpus* composto de 283 textos extraídos do jornal de Pediatria em português (http://www.jped.com.br/). Na comparação proposta nesse trabalho, a abordagem linguística obteve melhores resultados de precisão na extração segundo a comparação com uma lista de termos de referência desenvolvida de forma semimanual (projetos TEXTQUIM-TEXTECC-www.ufrgs.br/textecc).

Enquanto a extração através de uma abordagem estatística pode ser utilizada em qualquer tipo de texto e mesmo em qualquer língua, a extração com abordagens linguística tende a ser bastante dependente pelo menos da língua na qual os textos estão escritos. Um outro exemplo de trabalho científico de extração de termos através de uma abordagem linguística para o tratamento de linguagem científica é o trabalho de Bourigault que também extrai sintagmas nominais de bases textuais acadêmico científicas. Neste trabalho os termos são extraídos de um *corpus* de Direito composto por 12 códigos de lei da França (Bourigault et alli, 2005). Neste trabalho o objetivo é a construção de uma hierarquia dos termos para acelerar um motor de busca de informações em um site de acesso a textos jurídicos em francês.

Na verdade, a constante aplicação de técnicas de PLN, especificamente, extração de termos, sobre textos escritos em linguagem científica se explica pela grande necessidade de um tratamento automático destes textos para aplicações tão variadas como geração automática de glossários, construção de mecanismos de busca,

categorização de textos e até construção de ontologias. Textos científicos são usualmente descritos em uma linguagem objetiva, que possui como finalidade a circulação do saber e, portanto, seguindo determinados padrões linguísticos. Estas características fazem com que esse tipo de texto possa ter um tratamento computacional mais controlável.

O uso de PLN para o tratamento de linguagens científicas apresenta diversas perspectivas de aplicação e grandes desafios de pesquisa. Os avanços obtidos desde o surgimento do PLN foram enormes, porém o problema de compreensão automática de linguagem natural, como sempre acontece com as atividades humanas, é provavelmente impossível de ser completamente resolvido. No entanto, o tratamento de idiomas como o inglês, conta com uma diversidade de recursos e ferramentas. Avanços semelhantes precisam ser feitos para a língua portuguesa. Por isso, o campo de pesquisas em PLN para o tratamento de textos científicos em português mostra-se em constante desenvolvimento, e um campo de pesquisa bastante promissor.

## Referências Bibliográficas

AHO, A. V.; ULLMAN, J. D. *The Theory of Parsing, Translation, and Compiling*, Vol. 1, Parsing. Prentice Hall, 1972.

BANBERJEE, S.; PEDERSEN, T. The Design, Implementation, and Use of the Ngram Statistics Package. In: *Proceedings of the Fourth International Conference on Intelligent*

*Text Processing and Computational Linguistics*, Mexico City, 2003.

BAR-HILLEL, Y. The present status of automatic translation of languages, *Advances in Computers* 1(1): 91-163, 1960.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web, *Scientific American*, 284(5): 34-43, May 2001.

BOURIGAULT, D.; FABRE, C.; FRÉROT, C.; JACQUES, M.; OZDOWSKA, S. SYNTEX, analyseur syntaxique de corpus, *TALN*, Dourdan, 2005.

BUITELAAR, P.; CIMIANO, P.; MAGNINI, B. Ontology learning from text: An overview. In: Buitelaar, P.; Cimiano, P.; and Magnini, B. (Ed.). *Ontology Learning from Text: Methods, Evaluation and Applications*, v. 123 of Frontiers in Artificial Intelligence and Applications. IOS Press, 2005.

BRESNAN, J.; KAPLAN, M.R. Introduction: Grammars as mental representations of language. In: J. Bresnan and M.R. Kaplan (ed.), *The Mental Representation of Grammatical Relations*, Cambridge, Massachusetts, The MIT Press 1982.

CIMIANO, P. *Ontology learning and population from text: Algorithms, evaluation and applications.* Heidelberg: Springer-Verlag, 2006.

COLBY, K. M. Simulation of Belief Systems. Computer Models of Thought and Language. In: R. C. Schank and K. M. Colby. San Francisco, W.H. Freeman and Company: 251-286, 1973.

COLMERAUER, A. Total Precedence Relations. *Journal of the ACM*, 17(1): 14-30, 1970.

COLMERAUER, A. *Les Grammaires de Metamorphos*, Technical Report, Groupe d'Intelligence Artificielle, Marseille-Luminy, November, 1975.

COLMERAUER, A.; ROUSSEL, P. *The birth of Prolog History of Programming Languages*. ACM Press, 1996.

CHOMSKY, N. *Syntactic Structures*. The Hague: Mouton. Reprint. Berlin and New York, 1957.

CHOMSKY, N. Aspects of the theory of syntax. Cambridge, Mass.: MIT Press, 1965.

CRUBÉZYB, M.; ERIKSSONC, H.; NOYB, N. F.; TUB, S. W. The evolution of Protégé: an environment for knowledge-based systems development. *International Journal of Human-Computer Studies*, 58(1): 89-123, 2003.

EHRIG, M. *Ontology Alignment: Bridging the Semantic Gap*. University of Karlsruhe Germany, Springer, 2007

FILLMORE, C. J. *Lexical entries for verb*. Dordrecht, Holland: D. Reidel, 1968.

FORTUNA, B.; GROBELNIK, M.; MLADENIC, D. *OntoGen: Semi-automatic Ontology* Editor. HCI International, Beijing, July 2007.

GARSIDE, R., G. Leech and G. Sampson. *The Computational Analysis of English: A Corpus Based Approach*. London: Longman, 1987.

GENNARI, J. H.; MUSENB M A.; FERGERSONB, R. W.; GROSSOD, W. E.; GROSZ, B.; SIDNER, C. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12(3): 175-204, 1986.

GROSZ, B. The Representation and Use of Focus in a System for Understanding Dialogs. *IJCAI*, 67-76, 1977.

GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43 (5-6): 907-928, 1995.

HENDRIX, G, G.. Human Engineering for Applied Natural Language Processing. In: *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, Cambridge, Massachusetts, 1977.

JURAFSKY, D.; MARTIN, J. H. *Speech and Language Processing*, 2nd ed. Prentice-Hall, 2009.

KAY, P.; MCDANIEL, C. K. On the logic of variable rules. *Language and Society*, 8: 151-187. Cambridge University Press, 1979.

KUCERA, H.; FRANCIS, W. N. *Computational Analysis of Present-Day American English*. Providence: Brown University Press, 1967.

KURAMOTO, H. Nominal Groups: a New Purpose to Information Retrieval. DataGramaZero

*Revista de Ciência da Informação*, 3(1), 2002.

LIDDY, E. D. Natural Language Processing. In: *Encyclopedia of Library and Information Science*, 2nd ed. New York: Marcel Decker, Inc., 2003.

LOPES, L.; VIEIRA, R.; FINATTO, M. J.; ZANETTE, A.; MARTINS, D.; RIBEIRO Jr., L. C. Extração automática de termos compostos para construção de ontologias: um experimento na área da saúde. *Reciis-Revista Eletrônica de Comunicação Informação & Inovação em Saúde*. 3: 76-88, 2009a.

LOPES, L.; OLIVEIRA, L. H.; VIEIRA, R. *Análise Comparativa de Métodos de Extração de Termos: Abordagens Linguística e Estatística*. Relatório Técnico, n. 053, PPGCC, FACIN PUCRS, Porto Alegre, 2009b.

LOPES, L.; VIEIRA, R. *EXATOLP - Extrator automático de Termos para Ontologias em Língua Portuguesa*. Relatório Técnico, n. 054, PPGCC, FACIN PUCRS, Porto Alegre, 2009c.

MANN, W.C.; THOMPSON, S.A. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3): 243-281, 1988.

MANNING, C. D. SCHUTZE, H. *Foundations of statistical natural language processing*. Cambridge, Massachusetts: The MIT Press, 1999.

MARCUS, M. P.; SANTORINI, B.; MARCINKIEWICZ, M. A. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2): 313–330, 1993.

MCDONALD, D. D. Natural Language Generation as a Computational Problem. In: Brady, M. and Berwick, Bob, editors, *Computational Models of Discourse*, 209-265. MIT Press, 1983.

MCKEOWN, K. R. *Text generation: using discourse strategies and focus constraints to generate natural language text*, Cambridge University Press, New York, NY, 1985.

MCGUINNESS, D.L.; VAN HARMELEN, F. OWL web ontology language overview. World Wide Web Consortium (W3C) recommendation. http://www.w3.org/ Acesso em: 01 Agosto, 2009.

OKBC - Open Knowledge Base Connectivity Working Group *http://www.ai.sri.com/~okbc/* Acesso em: 01 Agosto, 2009.

PAPINENI, K.; ROUKOS, S.; WARD, T.; ZHU, W. BLEU: a method for automatic evaluation of machine translation. In: *Proceedings of the 40th Annual Meeting on Association For Computational Linguistics* (Philadelphia, Pennsylvania, July 07 - 12, 2002). Annual Meeting of the ACL. Association for Computational Linguistics, Morristown, NJ, 311-318, 2002.

PEREIRA, F. C. N. & WARREN, D. H. D. Definite Clause Grammars for Language Analysis – A Survey of the Formalism and a Comparison with Augmented Transition Networks. *Artificial Intelligence*, 13: 231-278, 1980.

QUILLIAN, M. Semantic Memory. In: M. Minsky, editor, *Semantic Information Processing*, 227-270, MIT Press, 1968.

RABINER, L. R., HUANG, B. H. An Introduction to Hidden Markov Models, *IEEE ASSP Magazine*, January 1986.

SCHANK, R. C. Conceptual Dependency: A Theory of Natural Language understanding, *Cognitive Psychology*, 3(4): 532-631, 1972.

SHANNON, C. *Communication Theory of Secrecy Systems*, 1949.

WALTZ, D. L. An English Language Question Answering System for a Large Relational Database, *Communications of the ACM*, 21(7): 526-539, 1978.

WEAVER, W. Translation. (reproduced in Locke and Booth, eds.: *Machine Translation of Languages,Cambridge*, MIT Press, 1955), 1949.

WEIZENBAUM, J. ELIZA - A computer program for the study of natural language communication between men and machines. *Communications of the ACM*, 9: 36-45, 1966.

WILKS, Y. Preference semantics. In: E. L. Keenan, editor, *Formal Semantics of Natural Language*. Cambridge University Press, 329-348, 1975.

WINOGRAD, T. *Procedures as a Representation for Data in a Computer program for Understanding Natural Language*. M.I.T. dissertation, 1971.

WITTEN, I.H; FRANK, E. *Data Mining: Practical machine learning tools and techniques*, 2 ed. San Francisco: Morgan Kaufmann, 2005.

WOODS, W. A. Transition Network Grammars for Natural language Analysis, *Communications of the ACM*, 13(10), 1970.

WOODS, W. A. Lunar rocks in natural english: Explorations in natural language question answering. In: A. Zampoli, editor, *Linguistic Structures Processing*, volume 5 of Fundamental Studies in Computer Science, 521-569. North-Holland, 1977.

# O VERBO MODAL *CAN* EM MANUAIS DE AVIAÇÃO EM INGLÊS: IMPLICAÇÕES PARA MATERIAIS DIDÁTICOS[1]

*Simone Sarmento* (UFRGS)

**Resumo**: *De forma quantitativa, e com base nos pressupostos da Linguística de Corpus, o presente trabalho se dedica a investigar a incidência dos Verbos Modais no segmento da linguagem da aviação representado pelo texto do tipo "manuais de aeronaves em inglês". Além disso, são observadas as colocações e os usos de "can", o modal mais frequente no corpus. Com base nos resultados da investigação, busco colher subsídios e oferecer sugestões para a elaboração de materiais didáticos em inglês com propósitos específicos*[2].

Este trabalho é fruto de uma longa trajetória de estudos e reflexões sobre a área da aviação, a respeito da qual sempre observei uma grande carência de pesquisas acadêmicas sistematizadas sobre as diversas sublínguas[3] que perfazem o total da área. No intuito de contribuir para suprir essa lacuna, este estudo se dedica a investigar mais profundamente o segmento da língua da aviação representado pelo gênero textual[4] "manuais de aeronaves em inglês". Com base nos

---

[1] O presente trabalho é baseado em Sarmento (2008).

[2] O presente trabalho é baseado na minha tese de doutorado intitulada "O uso dos verbos modais em manuais de aviação em inglês: Um estudo baseado em corpus", defendida no Programa de Pós-Graduação em Letras da UFRGS em 2008.

[3] Uma sublíngua é uma variedade da língua usada em certa ciência ou tecnologia que não somente é menor do que o total de uma língua, mas é também mais sistemática na sua estrutura e significados. É um subsistema que se comporta essencialmente como o total desta língua, ao mesmo tempo em que é limitado a um domínio específico. Cada sublíngua possui uma gramática e um léxico distinto (KITTREDGE, 1982). Outras sublínguas da área da aviação são a comunicação oral entre pilotos e controladores de tráfego aéreo, a comunicação via sistema de som entre pilotos e passageiros, entre muitas outras.

[4] Os gêneros são classes de eventos comunicativos que possuem características de estabilidade, reconhecimento de termos, etc. Para Swales (1990, p.46) eventos comunicativos compreendem não somente o próprio discurso e seus participantes, mas também o papel do discurso e seu ambiente de produção e recepção.

resultados, busco colher subsídios para elaborar materiais didáticos que reflitam a língua efetivamente usada nos manuais.

Pesquisas na área das línguas especializadas apontam para o fato de que

> Os textos técnicos são escritos baseados em certas suposições (ou pré-suposições) culturais relacionadas ao tipo e quantidade de informações gramático-retóricas que os redatores acreditam que os leitores compartilhem. [...] Nossa pesquisa mostra que grande parte dos falantes (leitores) não-nativos de inglês não possui esse *background* cultural que os habilite a ter essas informações pressupostas de forma a entender os textos técnicos[5]. (TRIMBLE, 1985, p. 114)

Dentre os elementos que causam maior dificuldade para alunos não-nativos de inglês, Trimble (1985, p. 115) afirma estar, em ordem decrescente, (a) distinções entre voz ativa e passiva em descrições e instruções; (b) uso de verbos modais em instruções; (c) uso não-padrão (e a ausência) de artigo definido em descrições e instruções; e (d) escolha de tempo verbal em descrições. O presente trabalho aborda de forma direta e/ou indireta os dois principais fatores causadores de dificuldades entre os usuários desse tipo de texto, ou seja, o foco principal deste estudo concentra-se na observação dos verbos modais em manuais de aviação. A ênfase secundária está na forma em que essas unidades se apresentam: voz ativa ou passiva. Investigações anteriores (SARMENTO, 2005a; SARMENTO, 2005b; GABRIELATOS e SARMENTO, 2006), mostram que verbos modais (doravante VMs) nos manuais de aviação são regularmente usados em orações passivas.

Os manuais em análise estão restritos à área da aviação, contudo, existem diferentes classes de profissionais envolvidos. Os manuais são escritos por técnicos redatores da BOEING[6] e serão lidos, sempre em inglês, por mecânicos e pilotos ao redor do mundo; há uma relação de "obediência" entre os interlocutores, tendo-se em mente que

---

[5] As tarduções das citações são de minha autoria.

[6] Não há um autor específico nos créditos dos manuais. A responsabilidade do texto, nesse caso, recai sobre a instituição como um todo.

os destinatários devem executar com precisão as diretrizes postas pelos redatores dos manuais. Pode-se dizer que os destinatários dos manuais aqui analisados formam duas comunidades discursivas (SWALES, 1990) distintas e ocupam diferentes posições na escala hierárquica de uma empresa aérea. Os pilotos ocupam a posição mais alta na empresa. O treinamento de um piloto é altamente dispendioso, custando às empresas aéreas alguns milhares de dólares ao ano. Os pilotos tornam-se assim, funcionários importantes de serem mantidos em uma empresa. Além disso, eles podem ocupar posições gerenciais e participar de importantes decisões em uma empresa, como por exemplo, decidir qual tipo de aeronave comprar. Os mecânicos, não possuem poder semelhante. Hierarquicamente, estão mais próximos à base. Outra diferença importante diz respeito ao local de trabalho. Para os pilotos, seguir as instruções em um manual não é só uma questão de eficiência, mas literalmente de "vida ou morte", uma vez que eles estão sempre a bordo das aeronaves. Os mecânicos realizam seu trabalho no solo. Sendo que uma ação erroneamente executada não colocará a sua vida em risco, mas sim dos passageiros e pilotos.

Os textos desses manuais são considerados portadores do conhecimento sobre as aeronaves, e não são passíveis de questionamentos por parte dos destinatários. Possuem um caráter altamente deôntico. São textos: descritivos, divididos em capítulos por partes da aeronave (hidráulica, trem de pouso, escadas, etc.) e/ou procedurais, divididos por tipo de procedimentos (decolagem, pouso, etc.). São ricos em informações não verbais, com várias tabelas, figuras e gráficos.

Assim, os manuais técnicos determinam, acima de tudo, um tipo de ação, expressando obrigações, possibilidades e consequências (SARMENTO, 2005a; SARMENTO, 2005b; GABRIELATOS e SARMENTO, 2006). Entre outras unidades lexicais, os VMs costumam ser usados para expressar essas funções comunicativas. São, assim, elementos lexicais fundamentais nesse cenário.

Com base nas reflexões expostas acima, este estudo tem por objetivo investigar o verbo modal *can* em dois manuais da aeronave BOEING 737. Busco descrever a incidência e a funcionalidade do VM nesse *corpus*. Realizo um contraste entre os manuais que compõem o

*corpus* e termino a análise com a investigação detalhada de *can* com o objetivo de oferecer subsídios para a elaboração de material didático voltado ao ensino de inglês para aviação tendo em vista seus diferentes propósitos.

Para dar conta dos objetivos, uso como base os pressupostos da Linguística de *Corpus* (doravante LC), das línguas especializadas e estudos sobre os VMs em inglês baseados em *corpora* ou com foco específico na aviação.

## 1. Verbos Modais

A modalização[7] é um assunto bastante complexo, sendo que em cada língua há formas diversas de expressar um mesmo tipo de modalidade (cf. OLIVEIRA, 2003 e COATES, 2005). Em contrapartida, uma mesma forma pode expressar diferentes modalidades, ou seja, os conceitos modais podem ser expressos nas línguas naturais através de uma grande variedade de formas.

Os VMs em inglês possuem características bastante específicas no que diz respeito à forma, ou seja, possuem um alto grau de gramaticalização[8]. Diferentemente de outras línguas, tal como a portuguesa, na qual os verbos modais têm marcação de concordância verbal e de tempo, os auxiliares modais da língua inglesa contemporânea não apresentam essas características. Para ilustrar as diferenças relacionadas à concordância verbal dos VMs entre português e inglês, cito o verbo modal "poder" em português, e "*can*" em inglês.

| Português | Inglês |
|---|---|
| Ele pode ir à festa. | *He can go to the party*. |
| Nós podemos ir à festa. | *We can go to the party*. |

Nos exemplos acima é possível notar que, em português, o VM *poder* concorda com os pronomes que o precede (*pode*, *podemos*). Já *can* mantém-se invariável independentemente do pronome. Não há

---

[7] Modalidade, em um sentido amplo, é relacionada às noções de "necessidade" e de "possibilidade" (LYONS, 1977).

[8] Gramaticalização é aqui entendida como *"*atribuição de um caráter gramatical a uma palavra anteriormente autônoma*"*(NEVES, 2004, p. 113).

essa variação em inglês nem mesmo para o pronome *he*, que, no presente simples dos demais verbos, receberia o morfema "*s*", como em *I work* e *He works*. Assim, tem-se *can* seguindo *he* e *we*.

| Português | Inglês |
|---|---|
| Eu posso ir à festa. | *I can go to the party.* |
| Eu pude ir à festa. | *I could go to the party*. |

Nesses últimos exemplos observa-se a marcação temporal relacionada a *posso* (presente do indicativo) e *pude* (pretérito) na língua portuguesa. Tal marcação não acontece em inglês, pois *can* foi utilizado no presente enquanto *could* foi utilizado no passado. De acordo com Denison (1993) e Bybee (1995), os modais em inglês são derivados de verbos que possuíam tempo e faziam concordância verbal durante um estágio bem anterior da língua. Dessa forma, os autores reconhecem que a derivação histórica desses verbos ainda traz algumas implicações semânticas como *can-could, will-would, may-might, shall-should*, em suas formas de presente e pretérito respectivamente. Apesar de haver essa derivação histórica de *could* como passado de *can*, esses dois itens lexicais são geralmente considerados independentes. Essa independência pode ser notada, por exemplo, em entradas lexicais nos dicionários. *Pude* não possui uma entrada própria em dicionários como o Aurélio ou o Houaiss, sendo apenas uma conjugação de verbo *poder*. Já *could* é uma entrada individual, totalmente independente de *can* no Macmillan English Dictionary.

Quanto à semântica dos VMs, faz-se necessário, primeiramente, distinguir dois tipos de modalidade: a epistêmica e a raiz. A primeira diz respeito às suposições do falante ou à avaliação de possibilidades. Na maioria dos casos, esse recurso indica a segurança (ou a sua ausência) que um falante possui acerca da verdade da proposição expressa (COATES, 1995), por exemplo, *They may be in the office* ou *They must be in the office* (PALMER, 2003, p. 7). A modalidade raiz engloba significados como permissão e obrigação, assim como possibilidade e necessidade e pode ser subdividida entre deôntica e dinâmica. A modalidade deôntica expressa uma permissão ou injunção do falante para que um ato seja realizado ou para que uma obrigação seja cumprida (HOYE, 1997). O evento é controlado por circunstâncias externas ao sujeito da oração

ou, em um sentido mais restrito, à pessoa ou às pessoas identificadas pelo sujeito, como nos exemplos *They may/can come in now* ou *They must come in now* (PALMER, 2003, p. 7). A dinâmica, por sua vez, trata da habilidade/capacidade ou disposição do sujeito da oração (PALMER, 1990); o controle é interno ao sujeito, vide exemplos: *They can run very fast* ou *I will help you*.

Mais recentemente, o fator distribucional,[9] motivado pelo avanço das pesquisas realizadas com o auxílio de computadores, tem recebido mais atenção, sendo que um dos mais importantes usos das investigações baseadas em *corpus* é precisamente "fornecer informações sobre a frequência do uso" (BIBER et al., 1999, p. 8). É importante ressaltar que, em consonância com os pressupostos da LC, a frequência de itens lexicais não é considerada aleatória, e sim motivada por necessidades comunicativas, entre outros fatores. Nesse sentido, torna-se útil que professores, acadêmicos, alunos e autores/produtores de textos saibam quais padrões gramaticais são regulares, merecedores, portanto, de mais atenção. Até muito recentemente, essa informação era geralmente baseada na intuição de falantes nativos de uma língua. Entretanto, as intuições dos falantes nativos são muitas vezes incorretas, ou pelo menos inexatas, pois

> cada um de nós possui apenas um conhecimento parcial da própria língua, temos nossas preferências e preconceitos, nossa memória é fraca, nossa imaginação é poderosa (assim podemos conceber contextos para as ocorrências mais implausíveis), e tendemos a prestar mais atenção nas palavras ou estruturas incomuns e não notar as mais comuns. (McENERY, XIAO & TONO, 2006, p. 147)

De forma a ilustrar a importância do fator distribucional, o Quadro 1 mostra as diferenças de frequência de ocorrência dos VMs em sete diferentes *corpora* conforme relatado na literatura da área (COATES, 1983; BIBER et al., 1999; KENNEDY, 2002; GABRIELATOS

---

[9] Por fator distribucional, refiro-me à regularidade de ocorrência dos VMs em diferentes gêneros textuais.

e SARMENTO, 2006; DERNARDIN, 2008)[10]. Os seis primeiros *corpora* apresentados no quadro são *corpora* de inglês geral. O OSC, na coluna de número sete, é um *corpus* de inglês específico composto de manuais de sistemas operacionais para computadores pessoais. Os VMs estão apresentados em ordem decrescente de ocorrência.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| **Frequência** | Survey | Lancaster | LSWE | FROWN | BNC | BROWN | OSC |
| ***1*** | *will* | *would* | *will* | *would* | *will* | *would* | *can* |
| ***2*** | *can* | *will* | *would* | *can* | *would* | *will* | *will* |
| ***3*** | *would* | *can* | *can* | *will* | *can* | *can* | *may* |
| ***4*** | *could* | *could* | *could* | *could* | *could* | *could* | *might* |
| ***5*** | *must* | *may* | *may* | *may* | *may* | *may* | *should* |
| ***6*** | *should* | *should* | *should* | *should* | *should* | *must* | *would* |
| ***7*** | *may* | *must* | *must* | *must* | *must* | *should* | *must* |
| ***8*** | *might* | *might* | *might* | *might* | *might* | *might* | *could* |
| ***9*** | *shall* | *shall* | *shall* | *shall* | *shall* | *shall* | *shall* |

Quadro 1: Classificação da frequência dos VMs entre diferentes *corpora*

Apesar de os seis *corpora* de inglês geral representarem diferentes variedades do inglês (americano e/ou britânico, oral e/ou escrito, geral ou específico), e terem sido compilados em diferentes épocas (dos anos 60 a 90), pode-se notar algumas similaridades com relação à frequência e distribuição dos VMs. Na primeira posição, encontram-se *will* e/ou *would*, sendo que *will* tem a preferência em quatro *corpora*. O modal *can* ocupa a segunda ou terceira posição nos

---

[10] O Survey utilizado por Coates (1983) continha 725 mil palavras de inglês escrito e oral. O LOB (Lancaster-Oslo/Bergen Corpus), também investigado por Coates possui um milhão de palavras de inglês escrito em 15 diferentes gêneros. O corpus LSWE (Longman Spoken and Written English Corpus) foi estudado por Biber Et al. (1999) contendo mais de 40 milhões de palavras. O FROWN, investigado por Gabrielatos e Sarmento, 2006) é um corpus de inglês americano escrito, composto de um milhão de palavras compilado nos anos 90. O BNC, analisado por Kennedy (2002), é um corpus de aproximadamente 100 milhões de palavras de IB escrito e falado, que foi disponibilizado em 1995. O BROWN consiste em um milhão de palavras de IA compilado no início dos anos 60. O OSC é um corpus de inglês escrito composto dos manuais dos sistemas operacionais para computadores pessoais: Windows e Linux. O primeiro consiste em 93.277 *tokens* e o segundo em 81.305, totalizando 174.582 *tokens*. O OSC (Operational System Corpus) foi compilado por Dernardin (2008) e estudado pelo mesmo autor juntamente com o BROWN.

*corpora* de inglês geral, sendo mais frequente na terceira posição. O *could* é o quarto modal mais frequente em todos os seis *corpora*. A quinta posição é ocupada por *may* em cinco e *must* em um dos *corpora*. Similarmente, a sexta posição é ocupada por *should* em seis *corpora* e *must* em um, enquanto *must* possui a sétima posição em quatro dos *corpora* investigados. Os modais *might* e *shall* ocupam respectivamente a oitava e a nona posições em todos os seis *corpora* de inglês geral.

Cotejando os dados dos *corpora* de inglês geral com o OSC, é possível observar algumas diferenças marcantes. O *can* é o VM mais frequente, diferentemente dos outros seis *corpora*. Em contrapartida, *would*, que nos outros *corpora* ocupa a primeira ou a segunda posição em ordem de frequência, é apenas o sexto no OSC. O mesmo pode ser dito sobre *could*, que é o quarto em todos os *corpora* de inglês geral sendo apenas o oitavo no OSC. *Should*, *must* e *shall* não apresentam maiores variações.

Além do fator frequência, a LC é também responsável pela investigação de aspectos que costumavam ser desconsiderados ou receber pouca atenção, como por exemplo, o estudo das colocações, neste trabalho entendidas como “a tendência com que certas palavras co-ocorrem com outras”. As listas de colocações podem ressaltar os diferentes significados de uma palavra, ou seja, os diferentes significados de uma palavra tenderão a ter diferentes colocados. Para que sejam definidos os padrões de uma palavra, faz-se necessário averiguar as palavras e as estruturas regularmente associadas a ela que de alguma forma refletem no seu significado.

## 2. O Modal *Can*

Apresento a seguir algumas breves características sobre os diferentes sentidos e usos do VM *can*. Há uma extensa variedade de perspectivas com relação ao tratamento teórico dos VMs. Entretanto, como o foco prioritário deste estudo é colher subsídios para a elaboração de material de inglês com propósitos específicos, é importante revisar o que a literatura voltada a professores de inglês (que não são necessariamente linguistas) oferece sobre o tema.

*Can* pode expressar possibilidade [*Even expert drivers* can *make mistakes*. (*It's possible for even*…)] (LEECH, 1989, p. 85). Leech afirma que esse sentido é muitas vezes encontrado na forma negativa com *cannot* ou *can't*: [*He* can't *be working at this hour*! (*It's not possible*…)]. Às vezes, *can* expressa um significado relacionado a hábitos, que pode ser parafraseado pelo advérbio *sometimes* [*Lightning* can *be very dangerous*. (*Lightning is sometimes very dangerous*.)] (LEECH, 1989, p. 86).

*Can* pode também expressar habilidade, com um sentido semelhante a *be capable of*, ou, quando for uma habilidade adquirida, com um sentido equivalente a *know how to* [*Paula* can't *sing, but she* can *play the guitar*. (*She knows how to*…) / *You* can *work harder than this*. (…*are capable of*…)] (LEECH, 1989, p. 86).

Segundo Leech (1989, p.86), com os verbos de "cognição inerte"[11] e de "percepção inerte"[12] há pouca diferença entre *being able to do something* e *actually doing something*. Em outras palavras, *I can remember* pouco difere de *I remember* como modo de se referir ao estado de lembrar. Com os verbos de percepção inerte, a principal função é expressar um estado (*I* can *hear*.) em contraste a uma percepção momentânea (*I hear*.).

Não há uma clara distinção entre os significados de possibilidade e de habilidade, uma vez que o último implica o primeiro. Uma possível marca distintiva para o significado de possibilidade é sua ocorrência em orações passivas [*This game can be played by young children*. (*It is possible for this game…)*]. Entretanto, a sentença na voz ativa [*(Even) young children* can *play this game*], pode também ser interpretada como habilidade.

Shawcross (1992), em seu livro que trata sobre a lingua técnica dos manuais de aviação, confirma a semelhança entre esses dois significados, apresentando os usos de *can* sob o título de possibilidade (SHAWCROSS, 1992, p. 93): "*Use Can + Infinitive to express possibility. (It's usually better than "it's possible to/that…".)*". Após essa breve explicação, Shawcross (1992, p. 94) oferece os seguintes exemplos:

---

[11] Essa categoria engloba verbos como *believe*, *forget*, *hope*, *imagine*, *know*, *suppose*, *understand* etc.
[12] Alguns exemplos de verbos assim classificados são *feel*, *hear*, *see*, *smell* e *taste*.

*The A320* ***can*** *carry 172 passengers.*

*Coaxial cables* ***can*** *transmit numerous messages.*

*In the event of an engine failure, the remaining engine* ***can*** *power all the hydraulic systems through the PTU.*

*The APU* ***can*** *be started up to 25,000 feet.*

O autor denomina o seguinte uso de *can* de “*capacity*”: “*The outer tank* ***can*** *hold 3,500kg*” (SHAWCROSS, 1992, p. 95)

Em língua geral, o *can* também é utilizado para pedir ou conceder permissão. [*You* can *stay here as long as you like. /* Can *I borrow your pen?*]

Conforme Leech (1989), linguistas mais tradicionais[13] consideram esse uso de *can* como “incorreto”. Entretanto, Coates (1983) e Leech (1989) sugerem que *can* é muito mais frequentemente utilizado como auxiliar de permissão do que *may*, forma tradicionalmente considerada correta. Coates (1983), Leech (1989) e Palmer (1990) sugerem que, nesses casos, *can* e *may* são praticamente intercambiáveis, sendo que a principal diferença é o fato de *may* ser considerado mais formal e, assim, mais cortês.

## 3. O *Corpus* de Estudo: *CORPUS* DA AVIAÇÃO

O *Corpus da Aviação* (doravante CA) compreende dois manuais da aeronave Boeing 737: (i) o manual de manutenção (MM) destinado a mecânicos e técnicos de aeronaves com um total de 249.691 palavras; e (ii) o manual de operações (MO) para pilotos, com 183.959 palavras. O MM[14] contém as informações necessárias para revisões, resoluções de problemas, verificações, reparos e substituições de todos os sistemas e equipamentos instalados nas aeronaves 737-300/400/500. Além disso, o MM contém informações relacionadas à

---

[13] Assim como pais e professores, que regularmente corrigem seus filhos e alunos.

[14] O MM não foi utilizado na íntegra, pois obtive acesso somente a algumas partes dele. Além disso, o manual como um todo é muito maior do que o manual para pilotos, somando 98 capítulos. Os capítulos utilizados foram 0 (*introduction*), 5 (*time limits/maintenance checks*), 12 (*servicing*) e 20 (*standard practices*).

inspeção e manutenção das estruturas das aeronaves. Seguem alguns exemplos de concordâncias do MM com o modal *can*.

| N | Concordance |
|---|---|
| 1 | THE OIL STAY ON YOUR SKIN. YOU CAN ABSORB POISONOUS |
| 2 | YOUR SKIN FOR A LONG TIME. YOU CAN ABSORB POISONOUS |
| 3 | OIL TOUCH YOUR SKIN. YOU CAN ABSORB TOXIC MATERIALS |
| 4 | OIL TOUCH YOUR SKIN. YOU CAN ABSORB TOXIC MATERIALS |
| 5 | SEALER FOR THE DECAL. YOU CAN BREAK THE EDGE SEAL WITH |
| 6 | USE AN EXTENSION HANDLE. YOU CAN CAUSE DAMAGE TO SEALED |
| 7 | THAN 55 PSI FUEL PRESSURE, YOU CAN . CAUSE DAMAGE TO THE |
| 8 | IF YOU ARE NOT CAREFUL YOU CAN CAUSE DAMAGE TO THE |
| 9 | IF YOU ARE NOT CAREFUL, YOU CAN CAUSE DAMAGE TO THE |
| 10 | FROM SURFACES THAT MOVE. YOU CAN CAUSE FAILURE OF MOVING |
| 11 | THE OTHER ENGINE PARTS. YOU CAN CAUSE INJURY TO YOURSELF IF |
| 12 | S 683-060 C_A_U_T_I_O_N_: IF YOU CAN DRAIN THE FUEL FROM THE |
| 13 | ONE THREAD . ENGAGED. YOU CAN COMPRESS THE UNION AND |
| 14 | THE VACUUM BREAK VALVE, YOU CAN DAMAGE THE VSCF. (4) ON |
| 15 | THE VACUUM BREAK VALVE, YOU CAN DAMAGE THE VSCF. (3) Remove |
| 16 | THE VACUUM BREAK VALVE, YOU CAN DAMAGE THE VSCF. (a) ON |

O MO fornece as informações relativas aos sistemas da aeronave, desempenho, procedimento e limitações nas operações, isto é, o que a tripulação de vôo necessita saber para operar a aeronave 737 de forma eficiente e segura durante as operações previstas. Seguem alguns exemplos de concordâncias do MO com o modal *can:*

| N | Concordance |
|---|---|
| 1 | extends outboard of measuring stick 4 can be used as an indication of the |
| 2 | Increased thrust lever activity can be expected when encountering |
| 3 | disabled. 8 Speed Brake Lever Actuator Can be armed to raise ground spoilers |
| 4 | enabled. 8 Speed Brake Lever Actuator Can be armed to raise ground spoilers |
| 5 | total hydraulic power failure the ailerons can be mechanically positioned by |
| 6 | realignment cycle before the airplane can be moved. April 28, 2000 737 |
| 7 | for the origin and destination airport can be selected for entry into the flight |
| 8 | –1000 feet. Airspeed Entry Airspeeds can be entered into the FMC as |
| 9 | windshear. Predictive windshear alerts can be enabled prior to takeoff by |
| 10 | windshear. Predictive windshear alerts can be enabled prior to takeoff by |

Na tabela 1, resumo as principais informações de cada manual, incluindo a sua dimensão.

Tabela 1: Síntese do CA

| Dados | Data | Destinatário | Objetivo | Tokens |
|---|---|---|---|---|
| **737-300/400/500 Manual de Manutenção (MM)** | Novembro / 1998 | Mecânicos de aeronaves | Informações necessárias para a manutenção das aeronaves nos hangares | **249.691** |
| **737-7I9 Manual de Operações (MO)** | Julho / 2000 | Pilotos | Informações relativas aos sistemas da aeronave; para ser utilizado em treinamento ou em vôo | **183.959** |
| ***Corpus* da Aviação (CA)** | | | | **433.650** |

## 4. Metodologia

"A observação contribui mais para a teoria do que a teoria contribui para a observação" (LEECH, 1992, p. 111). Seguindo nessa linha e diante da inexistência de um modelo específico para a análise de VMs em textos técnicos, procurei estabelecer os princípios que nortearam a escolha e a delimitação das unidades de análise, bem como a estruturação das fases da pesquisa. A metodologia tem o reflexo de estudos prévios, que contribuíram para a compreensão do emprego dos VMs nos manuais de aviação e para o amadurecimento do presente trabalho (SARMENTO, 2005a; SARMENTO, 2005b; SARMENTO 2007; GABRIELATOS e SARMENTO 2006).

Os manuais utilizados já estavam em formato digital, em pdf não protegido. Isso exigiu que os textos fossem transformados para arquivo.txt para que pudessem ser estudados através da ferramenta *WordSmith Tools*, pois a ferramenta somente aceita textos no formato .txt. Após a conversão, foi feita uma *limpeza* manual dos textos. O *WordSmith Tools* é um conjunto de programas integrados destinados à análise linguística. O programa foi desenvolvido em 1996 por Mike Scott e publicado pela Oxford University Press e está atualmente em sua quinta versão. Neste trabalho faço uso da versão 5[15]. O *WordSmith*

[15] A licença para utilizar tal ferramenta pode ser obtida através do site: http://www.lexically.net/wordsmith/.

possui as ferramentas- *WordList, Keywords, Concord*. Neste trabalho utilizei principalmente algumas funções do *WordList* e do *Concord*.

Num primeiro momento utilizei o *WordList* para verificar o número exato de *tokens* em cada um dos subcorpora que compõe o CA. Após esse primeiro levantamento, foi utilizado o seguinte roteiro metodológico:

i) Verificação das diferenças quanto à frequência do VM *can* e de cada estrutura em cada um dos dois subcorpora que compõem o *corpus* de estudo.

Foi utilizada a ferramenta *Concord* para obter as concordâncias através de uma lista contendo todas as ocorrências de *can* e seu cotexto. Cada palavra investigada é destacada no centro de um fragmento do texto.

Logo após, para levantamento das estruturas sintáticas, utilizei como ponto de partida as nove estruturas gramaticais sugeridas por Kennedy (2002), conforme o quadro abaixo:

| | Estrutura | Exemplo |
|---|---|---|
| **1** | Modal sozinho | **(Who will go?) I will.** |
| **2** | Modal + infinitivo<br>(modal + voz ativa) | **Sam *can* swim./ She *must* be hungry./<br>You *should* have a rest.** |
| **3** | Modal + *be* + passado particípio<br>(modal+ voz passiva) | **It *should* be replaced.** |
| **4** | Modal+*be*+presente particípio<br>(modal+progressivo) | **They *will* be arriving soon.** |
| **5** | Modal+*have*+passado particípio (ou adjetivo)<br>(modal+perfeito) | **He *might* have done it.<br>He *must* have been hungry.** |
| **6** | Modal+*be*+*being*+passado particípio (ou adjetivo)<br>(modal+be+progressivo+passado particípio/adjetivo) | **It *might* be being done tomorrow.<br>He *could* be being awkward.** |
| **7** | Modal+*have*+*been*+passado particípio<br>(Modal+voz passiva+perfeito) | **It *should* have been fixed.** |
| **8** | Modal+*have*+*been*+presente particípio<br>(modal+voz passiva+progressivo) | **He *must* have been lying.** |
| **9** | **Modal+*have*+*been*+*being*+passado particípio<br>(ou adjetivo)<br>(modal+voz passiva+perfeito+progressivo)** | **He *might* have been being blackmailed.<br>They *must* have been being careless.** |

Quadro2: Estruturas sintáticas dos modais
(KENNEDY, 2002, p. 82)

Foi utilizado o *Concord*, do *WordSmith* e feita uma análise manual de todas as linhas de concordância.

Primeiramente, as linhas de concordância foram analisadas com vistas a classificá-las conforme as nove estruturas presentes na tabela proposta por Kennedy (2002). As linhas foram salvas em arquivos separados já classificadas nessas estruturas, pois, a partir deste ponto, as outras análises serão sempre feitas conforme as estruturas em que *can* se encontra.

ii) Identifico as colocações de *can* em cada um dos subcorpora.

Os resultados mostram que no *Corpus* de Aviação (CA), há duas estruturas sintáticas predominantes, conforme a tabela proposta por Kennedy (2002):

a) Estrutura 2- modal +infinitivo (modal + voz ativa)
   Exemplo:
   *Airplanes* ***may*** *operate with uncured fay sealant or aerodynamic sealant on wing external surface.*
b) Estrutura 3- modal + *be*+ passado particípio (modal+ voz passiva)
   Exemplo:
   *The following is a list of cleaners and polishes that* ***could be*** *required . during regular maintenance of the airplane*.

A lista de colocados só pode ser obtida se alguns critérios forem cumpridos. Um desses critérios é que, pelo menos um dos colocados do nódulo apresente uma frequência mínima de quatro unidades em um horizonte de seis palavras, três à esquerda (L1, L2 e L3) e três à direita (R1, R2 e R3). Os resultados serão apresentados da seguinte forma:

As duas sub-etapas são:

A: Colocados à esquerda

B: Colocados à direita

Descrevo e explico agora essas duas sub-etapas e mostro as categorias que foram usadas para classificar as ocorrências.

A: Colocados à esquerda (L1, L2, L3)

Relação dos 10 colocados à esquerda com maior grau de atração.

Uma lista de colocados fornece uma espécie de perfil semântico das palavras envolvidas. Essas listas de colocados foram feitas para

cada modal em cada uma das estruturas encontradas . Para medir a força ou o grau de cada colocação, foi utilizado o teste estatístico *Log Likelihood.* Se o resultado obtido após a aplicação do LL foi de 6,63 ou mais, a probabilidade de a atração entre os dois itens lexicais ter acontecido aleatoriamente é de menos de 1%. Dessa forma, o pesquisador pode estar 99% certo de que o resultado é significativo. A função da estatística é apontar se os itens realmente formam colocações ou se são co-ocorrências espúrias. Os cálculos não foram realizados por mim, mas pelo próprio *WordSmith*, que apresenta também outras opções de testes estatísticos para medir a “força” de uma colocação (HUNSTON, 2002).

Mede-se o grau de associação entre duas palavras tendo em vista o número de vezes que essas duas palavras ocorrem juntas. Ou seja, quanto mais vezes dois elementos ocorrem simultaneamente e menos vezes ocorrem separadamente, mais forte será sua associação. Um elevado valor da medida de associação indica que as variáveis em causa estão associadas, isto é, a ocorrência de uma delas está associada à ocorrência da outra. Cada estatística tende a trazer resultados diferentes. De modo sucinto, a diferença entre as estatísticas é que, alguns testes (por exemplo, o *Mutual Information*, MI) capturam associações em que os itens aparecem juntos frequentemente em termos relativos, enquanto outros apontam associações em que os itens co-ocorrem frequentemente em termos absolutos. Isso significa dizer que o MI localiza pares de palavras que aparecem juntos muitas vezes em relação à frequência de um e de outro, mesmo que cada um ocorra poucas vezes. Já o LL dá preferência aos colocados que são, por si só, muito frequentes. Como o objetivo deste trabalho é descrever o que é mais frequente no CA para auxiliar na elaboração de material didático, entendo o LL mais apropriado para esse fim.

Dessa forma, são considerados colocados, os itens lexicais que se coloquem dentro de um horizonte de três palavras à esquerda (posições L1, L2 e L3) a três palavras à direita (posições R1, R2 e R3) do nódulo, com uma frequência mínima de quatro ocorrências, e uma relação LL de pelo menos 6,6. Através da lista de colocados é possível verificar se algum item lexical se destaca em termos de frequência ou

força de associação. Assim, nessa fase, já é possível verificar algumas características relativas ao comportamento do nódulo no *corpus*. A partir da verificação das listas de colocados à esquerda do nódulo, foi possível observar uma variação na classe de palavras que ocupam a posição L1, isto é, uma posição à esquerda do nódulo, que corresponderia teoricamente ao sujeito do VM.

Foi necessário analisar todas as linhas de concordância do nódulo *can* de forma a verificar que tipo de estrutura compunha essas posições à esquerda (L1, L2, L3). A partir dessa verificação, foi possível classificar os colocados à esquerda em cinco grupos: (1) *You*; (2) anáforas; (3) unidades simples; (4) sintagmas nominais; (5) sintagmas verbais e preposicionais. Essas categorias surgiram dos dados e não foram baseadas em uma taxonomia pré-existente. Como este trabalho visa fornecer subsídios para a elaboração de material didático, considerei importante classificar os tipos de elementos lexicais que antecedem o VM *can* por acreditar que alguns deles impõem mais dificuldade na compreensão da função do modal, uma vez que apresentam o sujeito separado do VM por outros elementos lexicais intermediários. Esse tipo de construção deve merecer mais destaque nos materiais didáticos. O único dos grupos (nas posições L1, L2, L3) constituído de um único item lexical é o grupo "*you*". Esses grupos são descritos a seguir:

(1) *You - You* não foi classificado em um grupo juntamente com outros pronomes pessoais retos por apresentar uma função diferente nesses textos. Por tratar-se de um tipo de texto com características deônticas que estabelece o que pode ou não ser feito, e o que é ou não possível, *you* difere dos usos dos outros pronomes pessoais que nesses textos, ocupam a função de um elemento anafórico. Segue alguns exemplos ilustrativos do *you* + *can*:

| N | Concordance |
|---|---|
| 1 | will dry on the surface before you can flush it with water. EFFECTIVITY□. |
| 2 | surface can dry the cleaners before you can flush them with water. The dried |
| 3 | it has these properties: (1) You can fill the tanks with fuel more quickly |
| 4 | IF THE OIL TANK IS COOL, YOU CAN FILL IT TOO MUCH AND CAUSE |
| 5 | of the tanks at the same time. (3) You can fill only one tank if it is necessary. |
| 6 | as far as the attached hardware. You can fully seal the attached hardware. If |
| 7 | of the fuselage, section 48 that you can get access to. Look at the structure |

(2) Anáforas - Neste grupo há sempre um elemento anafórico na posição L1, como por exemplo, *this, which, it, that, they, these* e *who.* Lembrando que elementos anafóricos são aqueles que retomam um elemento referencial**.** Vide os exemplos abaixo:

| N | Concordance |
|---|---|
| 25 | beam. N_O_T_E_: A device which can find mercury (mercury sniffer) is |
| 26 | movement of items or equipment which can hit the airplane or injure persons. |
| 27 | bearings, screens, screws, etc. This can also include capping of lines, |
| 28 | in order to perform the task. This can also include uncapping of lines, |
| 29 | TEMPERATURES CAN OCCUR. THIS CAN CAUSE A FIRE OR EXPLOSION |

(3) Unidades simples - Foram consideradas unidades simples os casos em que há apenas um substantivo, precedido por um determinante ou não. Seguem alguns exemplos ilustrativos deste grupo:

| N | Concordance |
|---|---|
| 2 | with hot or cold water immediately. Alcohol can also be used. S 162-036 (5) Dry the |
| 3 | BEFORE YOU START THE APU. THE APU CAN BE DAMAGED BY THE ICE OR SNOW |
| 4 | Strike Zone (Zone 1, Fig. 201). Damage can also be found on the skin trailing edge |
| 5 | the damaged area you can see. Delamination can be found by instrumental NDI procedures |
| 6 | TIRE OR WHEEL. AN EXPLOSION CAN BE CAUSED AND INJURY TO |

(4) Sintagmas nominais - Foram incluídos nesse grupo os colocados à esquerda representados por sintagmas nominais com pelo menos dois elementos, como mostra os exemplos abaixo:

| N | Concordance |
|---|---|
| 1 | show more oil than there is. Expanded air can push the oil from the oil cooler into the |
| 2 | (450¡F). Heat guns and soldering guns also can contain electrical switches which can |
| 3 | HIGH HAZARDOUS GAS CONTENT AREA CAN CAUSE AN EXPLOSION. THE |
| 4 | the airplane. C. Many small leakage areas can add to create an excessive fuselage |
| 5 | W_A_R_N_I_N_G_: VOLCANIC ASH CAN IRRITATE YOU EYES AND BE A |
| 6 | power in the battery. A fully charged battery can operate the refuel system for 15 to 20 |

(5) Sintagmas verbais e preposicionais - Este grupo é constituído de dois tipos de colocados à esquerda: os sintagmas verbais, por exemplo: *oil system parts that are hot can;* e sintagmas preposicionais, por exemplo, *damage to the valve can occur.* A característica em comum neste grupo é o fato de o sujeito, ou referencial, do VM não estar diretamente (imediatamente) ligado a ele, conforme os exemplos abaixo:

| N | Concordance |
|---|---|
| 1 | impedance changes from 10mp to 100mp can possibly be a cable connector and are |
| 2 | TO THE UNIT YOU ARE NOT SURE ABOUT CAN MELT AND LET SOME WATER AND |
| 3 | of the external surfaces of the airplane can be below freezing, ice can occur and |
| 4 | LOAD LIMITS. DAMAGE TO THE AIRPLANE CAN OCCUR. (a) Calculate the value that each |
| 5 | AND SLATS ARE FAST ACTING AND CAN CAUSE SERIOUS INJURY TO |
| 6 | OIL WILL PUT STAINS ON CLOTHES AND CAN MAKE PAINT SOFT. (3) Remove the filler |

É importante ressaltar que, para a classificação dos grupos "*You*", "anáforas" e "unidades simples" foram considerados somente os colocados na posição L1. Para os "sintagmas nominais" e "sintagmas verbais e preposicionais", não foi estabelecido uma posição limite, uma vez que o sujeito pode ser um composto nominal de até quatro elementos, ou uma oração completa. Outro fato a ser comentado é que com exceção do grupo "*you*", os outros grupos são compostos por elementos inanimados na posição de sujeito.

B: Colocados à direita (R1, R2, R3)

Analiso os colocados à direita de *can* em cada uma das estruturas a partir da lista de colocados. Mostro os 10 colocados com maior grau de atração ao nódulo (quando houver 10). Diferentemente dos colocados nas posições à esquerda, que podem pertencer a diferentes classes gramaticais, a posição R1 apresenta uma restrição

gramatical quanto ao tipo de elemento lexical que pode seguir o VM. Dessa forma, o VM pode ser imediatamente seguido apenas por verbo ou advérbio. Nos casos de VM + voz passiva, o *be* não é considerado um colocado do VM por ser considerado elemento obrigatório para estabelecer a estrutura. Assim, nos casos de voz passiva, a posição R1 é a primeira posição após o *be*.

Finalmente, com o objetivo verificar o sentido de cada ocorrência, são usados os resultados das etapas anteriores. Parto das categorias e exemplos apresentados no referencial teórico para analisar os sentidos, usos e peculiaridades de *can* no CA. Entretanto, também verifico a possibilidade de o VM estar expressando uma função não antes descrita. Apresento exemplos retirados do *corpus* que expressam os usos mais salientes em cada um dos subcorpora. Foi observado um contexto estendido, além de somente a linha de concordância, sempre que necessário para classificar o uso expresso pelo VM.

## 5. *Can* e suas companhias

Nesta seção, os resultados serão apresentados separadamente por manual, uma vez que são escritos para públicos alvo diferentes. Apresentarei os usos e funções do VM *can* nos manuais e mostrarei as ocorrências mais significativos desse VM para fins didáticos, ou seja, os exemplos que poderiam constar em materiais elaborados para mecânicos e pilotos.

### *Can* no Manual de Manutenção (MM)

No MM, o modal *can*, apresenta um total de 809 ocorrências, sendo 723 ocorrências na voz ativa e 86 ocorrências na voz passiva.

*<u>Colocados à esquerda</u>*

A Tabela 2 mostra os dez principais colocados à esquerda de *can* juntamente com a quantidade bruta de ocorrências.

Tabela 2: Colocados à esquerda de *can* no MM

| Posição | Voz ativa | | | Voz passiva | | |
|---|---|---|---|---|---|---|
| | Colocado | Quantidade | Relação LL | Colocado | Quantidade | Relação LL |
| **1** | you (L1) | 190 | 1.292,39 | sealed (L2) | 4 | 25,70 |
| **2** | which (L1) | 27 | 190,39 | force (L3) | 4 | 23,86 |
| **3** | this (L1) | 34 | 190,31 | they (L1) | 4 | 16,75 |
| **4** | that (L1) | 37 | 179,18 | bearings (L1) | 4 | 11,00 |
| **5** | oil (L1) | 33 | 168,68 | grease (L3, L2) | 4 | 8,60 |
| **6** | it (L1) | 26 | 152,53 | which (L1) | 4 | 7,65 |
| **7** | agents (L2) | 12 | 109,64 | ------ | ------ | ------ |
| **8** | hot (L1) | 16 | 109,62 | ------ | ------ | ------ |
| **9** | explosion (L1) | 9 | 106,88 | ------ | ------ | ------ |
| **10** | flammable (L3) | 12 | 103,76 | ------ | ------ | ------ |

Destaca-se sobremaneira a frequência e a força de atração de *you can* na voz ativa. É também surpreendente, a variedade de diferentes classes de palavras que ocupam a primeira posição à esquerda de *can*, com vários determinantes (*which, this*, e *that*), e um adjetivo (*hot*). Por tratar-se de um verbo, a expectativa era de haver substantivos, como no caso de *oil* e *explosion* ou pronomes pessoais retos, como *you*.

A lista de colocados fornece pistas a respeito do nódulo. Entretanto, faz-se necessário conhecer e classificar todos os colocados à esquerda de *can*. Após a análise de todas as linhas de concordância de *can*, os resultados foram agrupados em cinco categorias, como mostra a Tabela 3.

Tabela 3: Classificação das unidades lexicais à esquerda de *can* no MM

| Voz ativa | | Voz passiva | |
|---|---|---|---|
| **Colocados** | Quantidade | Colocados | Quantidade |
| ***You* (L1)** | 190 / 26,3% | *You* (L1) | ----- |
| **Anáforas (L1)** | 97 / 13,4% | Anáforas (L1) | 13 / 15,1% |
| **Unidades simples (L1)** | 88 / 12,2% | Unidades simples (L1) | 15 / 17,4% |
| **Sintagmas nominais** | 149 / 20,6 % | Sintagmas nominais | 28 / 32,6% |
| **Sintag. verbais e prepos.** | 199 / 27,5% | Sintag. verbais e prepos. | 30 / 34,9% |
| **Total** | 723 / 100% | Total | 86 / 100% |

Como pode ser visto, o *you* sozinho responde por 26,3% das ocorrências de *can* na voz ativa no MM. Os outros grupos são formados por variados itens lexicais. Como era de se esperar, o *you* não está presente nos casos de *can* +voz passiva, pois os colocados à esquerda em uma oração passiva não são geralmente pronomes pessoais retos.

*Colocados à direita*

Mostro agora os colocados à direita de *can*.

Tabela 4: Colocados à direita de *can* no MM

| Posição | Voz ativa | | | Voz passiva | | |
|---|---|---|---|---|---|---|
| | Colocado | Quantidade | Relação LL | Colocado | Quantidade | Relação LL |
| **1** | cause (R1) | 218 | 2.197,82 | damaged (R1) | 11 | 56,66 |
| **2** | occur (R1) | 84 | 844,07 | used (R1) | 12 | 53,63 |
| **3** | damage (R2) | 72 | 702,85 | by (R2) | 13 | 45,65 |
| **4** | injury (R2) | 43 | 535,31 | found (R1) | 6 | 30,21 |
| **5** | use (R1) | 52 | 239,65 | that (R2) | 4 | 7,60 |
| **6** | be (R1) | 46 | 225,43 | with (R2) | 7 | 6,79 |
| **7** | injuries (R2) | 15 | 150,42 | ------ | ------ | ------ |
| **8** | come (R1) | 12 | 139,47 | ------ | ------ | ------ |
| **9** | get (R1) | 26 | 135,12 | ------ | ------ | ------ |
| **10** | burn (R1) | 13 | 134,38 | ------ | ------ | ------ |

Nesta lista dos principais colocados à direita de *can* no MM, nota-se que aparecem apenas verbos na posição R1, como esperado. É bastante saliente, entretanto, a frequência e atração do verbo lexical *cause* na voz ativa. As 218 ocorrências de *can cause* correspondem a mais de 30% do total de ocorrências de *can* (voz ativa) no MM. Os verbos *occur* e *use* também apresentam uma forte atração com o modal *can*. Destaco que o cópula *be,* entre os colocados de *can* na voz ativa, não é um verbo auxiliar como no caso das construções passivas, mas o verbo principal. Ele é o quarto verbo com maior atração a *can*.

Nas ocorrências de *can* na voz passiva, os verbos com maior grau de atração são *damaged* e *used*, que apresentam um grau de atração bem inferior a *cause* na voz ativa. A ser destacado, mas que na verdade não apresenta qualquer surpresa, é a presença da preposição *by*, que é geralmente usada em construções passivas para apontar a

entidade que realizou a ação. Além disso, destaca-se também a presença do verbo *found*.

<u>Usos e funções</u>

Após a análise detalhada de todas as ocorrências, foi constatado que no MM:

- *Can* é usado com maior frequência na voz ativa, na proporção de nove ocorrências na voz ativa para apenas uma na voz passiva.
- É usado principalmente na forma afirmativa, na proporção de uma ocorrência negativa (seguida pelo advérbio *not*) a cada dez afirmativas.
- Todas as ocorrências afirmativas de *can* expressam a idéia de possibilidade, podendo essa possibilidade ser relacionada a uma prosódia semântica[16] neutra ou negativa
- Quando *can* é acompanhado por um verbo causativo[17] (*cause*, por exemplo) ou de ocorrência (*occur*, por exemplo), ele é utilizado para expressar uma consequência negativa que pode acontecer devido a uma ação mal realizada pelo mecânico. Em outras palavras, nesses casos, *can* possui uma prosódia semântica negativa.
- *Can* aparece frequentemente nas seções de *warning* (aviso) ou *caution* (cuidado), com um total de 358 ocorrências. O uso de *can* nessas seções é sempre vinculado a uma prosódia semântica negativa.
- *Can* acompanhado de outros tipos de verbos não apresenta a característica de expressar uma consequência negativa de forma tão marcante. Sua função geralmente expressa apenas uma possibilidade mais neutra, mostrando possíveis ações a serem realizadas.

---

[16] Prosódia semântica é o termo usado para referir a palavra ou expressão usada em um contexto específico de tal forma que a palavra/expressão adquira certa conotação naquele contexto. Um exemplo seria a expressão *sit through* (HUNSTON, 2002, p. 141), que é geralmente usada com itens que indicam algo longo e cansativo, ou seja, assume uma conotação de algo tedioso, portanto uma conotação negativa.

[17] Taxonomia utilizada por Biber et al (1999).

- *Can* ocorre com certa frequência em orações condicionais com a presença da conjunção *if* (74 ocorrências), principalmente na oração principal de uma condicional (60 ocorrências). Entretanto, *can* também aparece na oração subordinada (14 ocorrências).
- Nos contextos negativos também predomina a função de possibilidade, ou melhor, de ausência de possibilidade devido a uma causa externa.
- Apenas três ocorrências de *cannot* expressam a idéia de forte recomendação negativa; isto é, recomendando ao mecânico não usar certos dispositivos, por exemplo:

C_A_U_T_I_O_N_: SOME OIL TYPES **CANNOT BE** MIXED WITH OTHER TYPES OF OIL. DO NOT MIX OIL TYPES, UNLESS YOU ARE SURE THAT THEY CAN BE MIXED OR ENGINE DAMAGE CAN OCCUR.

Com base nos resultados apresentados, ilustro as características e usos mais salientes de can no manual de manutenção.

1-C_A_U_T_I_O_N_: CAREFULLY MOVE ROPES OR FABRIC HOSES ON THE WING OR FUSELAGE. EQUIPMENT THAT IS INSTALLED ON THE SURFACE OF THE WING OR FUSELAGE **CAN** BE DAMAGED BY THE MOVEMENT OF THE ROPES OR FABRIC.

2- N_O_T_E_: If the fan (N1) **cannot** turn during an engine start because of ice, this **can** cause heavy damage.

3- If you **can** see the lining after bolt removal, refer to the bench check instructions and examine the lining material.

4- W_A_R_N_I_N_G_: INSTALL LEADING EDGE FLAP AND SLAT LOCKS TO PREVENT INADVERTENT OPERATION OF LEADING EDGE FLAPS AND SLATS. FLAPS AND SLATS ARE FAST ACTING AND **CAN** CAUSE SERIOUS INJURY TO PERSONNEL.

5- Immediately after refueling, the fuel quantity shown on the FMC **can** be less than the actual fuel quantity in the tank.

**Can no Manual de Operações (MO)**

No MO, can é o modal mais frequente apresentando um total de 327 ocorrências, sendo 85 ocorrências na voz ativa e 242 ocorrências na voz passiva.

Colocados à esquerda

A Tabela 5 mostra os dez principais colocados à esquerda juntamente com a quantidade bruta de ocorrências.

Tabela 5: Colocados à esquerda de *can* no MO

| Posição | Voz ativa | | | Voz passiva | | |
|---|---|---|---|---|---|---|
| | Colocado | Quantidade | Relação LL | Colocado | Quantidade | Relação LL |
| **1** | it (L1) | 4 | 19,42 | but (L2) | 4 | 41,34 |
| **2** | FMC (L3) | 5 | 8,33 | mode (L1) | 12 | 36,83 |
| **3** | APU (L1) | 4 | 7,50 | thrust (L2) | 10 | 36,46 |
| **4** | ------- | ------- | ------- | modes (L1) | 7 | 36,41 |
| **5** | ------- | ------- | ------- | it (L1) | 6 | 34,04 |
| **6** | ------- | ------- | ------- | waypoints (L1) | 6 | 31,78 |
| **7** | ------- | ------- | ------- | entry (L1) | 8 | 31,39 |
| **8** | ------- | ------- | ------- | this (L2) | 7 | 30,75 |
| **9** | ------- | ------- | ------- | information (L1) | 7 | 30,35 |
| **10** | ------- | ------- | ------- | executed (L2) | 4 | 28,59 |

Não há um item lexical que apresente uma preferência que se destaque. Nota-se, por outro lado, a ausência do pronome *you* e a baixa ocorrência de elementos anafóricos.

Tabela 6: Classificação das unidades lexicais à esquerda de can no MO

| Voz ativa | | Voz passiva | |
|---|---|---|---|
| **Colocados** | Quantidade | Colocados | Quantidade |
| ***You* (L1)** | ------- | *You* (L1) | ------- |
| **Anáforas (L1)** | 4 / 4,7% | Anáforas (L1) | 13 / 5,4% |
| **Simples (L1)** | 21 / 24,7% | Simples (L1) | 54 / 22,3% |
| **Sintagmas nominais** | 30 / 35,3% | Sintagmas nominais | 126 / 52,1% |
| **Sintag. verbais e prepos.** | 30 / 35,3% | Sintag. verbais e prepos. | 49 / 20,2% |
| **Total** | 85 / 100% | Total | 242 / 100% |

Como pode ser visto, o grupo dos sintagmas nominais é preferido por *can* na voz ativa e empata com o grupo de sintagmas verbais e preposicionais na voz passiva. Os elementos anafóricos apresentam poucas ocorrências nas duas estruturas.

Colocados à direita

Mostro agora os colocados à direita de *can*.

Tabela 7: Colocados à direita de *can* no MO

| **Posição** | **Voz ativa** | | | **Voz passiva** | | |
|---|---|---|---|---|---|---|
| | Colocado | Quantidade | Relação LL | Colocado | Quantidade | Relação LL |
| **1** | cause (R1) | 8 | 61,130 | selected (R1) | 20 | 128,946 |
| **2** | include (R1) | 7 | 58,679 | by (R2) | 34 | 125,537 |
| **3** | supply (R1) | 7 | 49,692 | used (R1) | 17 | 108,128 |
| **4** | also (R1) | 7 | 39,925 | reengaged (R1) | 5 | 59,708 |
| **5** | power (R2) | 7 | 32,594 | also (R1) | 9 | 56,011 |
| **6** | operate (R1) | 5 | 32,329 | from (R2) | 12 | 55,289 |
| **7** | occur (R1) | 5 | 28,880 | opened (R1) | 5 | 54,536 |
| **8** | provide (R1) | 5 | 28,313 | on (R2) | 19 | 51,752 |
| **9** | either (R2) | 4 | 20,890 | made (R1) | 7 | 49,970 |
| **10** | be (R1) | 7 | 17,651 | engaged (R1) | 6 | 47,499 |

Nesta lista dos principais colocados à direita de *can* no MO, nota-se que os verbos que ocorrem na voz ativa não se repetem na voz passiva, sendo *cause* o mais frequente na voz ativa e *selected* na passiva. A preposição *by* é um colocado forte de *can* (voz passiva), ocorrendo em aproximadamente 11% dos casos. O advérbio *also* é o único item lexical que aparece nas duas estruturas.

<u>Usos e funções</u>

Foi constatado que:

- *Can* é usado com maior frequência na voz passiva, na proporção de três ocorrências na voz passiva para uma ocorrência na voz ativa.
- É usado principalmente na forma afirmativa, na proporção de uma ocorrência negativa a cada cinco afirmativas.
- Todas as ocorrências de *can* em frases afirmativas expressam possibilidade.
- *Can*, quando acompanhado por um verbo causativo ou de ocorrência expressa uma consequência negativa que pode acontecer devido a uma ação mal realizada pelo piloto ou pelo mau funcionamento da aeronave.
- A função de *by* nas ocorrências de *can* na voz passiva indica um agente animado, neste caso, pilotos (*flight crew*) em apenas quatro casos, ou 12%. Nos outros casos, *by*

indica o procedimento que pode gerar a ação já descrita à esquerda de *can*.

Com base nos resultados apresentados acima, ilustro as características e usos mais salientes de *can* no MO.

*1-The engine indications* ***can*** *also be manually selected to either the Captain's or First Officer's inboard DU, or the lower DU, using the respective display select panel.*

*2- Air for the bleed air system* ***can*** *be supplied by the engines, APU, or an external air cart/source..*

*3- After localizer and glideslope capture during a dual autopilot approach, CWS* ***cannot*** *be engaged by manually overriding pitch and roll..*

*4- A spinning tire with a loose tread must be stopped prior to entering the wheel well or it can* ***cause*** *damage to wheel well components.*

## 6. Considerações Finais

Foi observada uma grande diferença na quantidade de colocados em cada um dos subcorpora na voz ativa. Os colocados[18] no MM apresentam uma frequência, ou força de colocação bem superior ao MO: por exemplo, o pronome *you* apresenta 191 ocorrências como colocado de *can*, ao passo que o colocado mais forte de *can* no MO apresenta somente 12 ocorrências (*mode*).

De forma geral, conclui-se que *can* encontra-se frequentemente em estruturas sintáticas bastante complexas. Essa constatação é relevante tendo em vista o objetivo final desta pesquisa que é o de obter subsídios para a elaboração de materiais didáticos para fins específicos. Não é suficiente apenas apresentar o VM e apontar o seu valor semântico, isto é, se o modal expressa uma possibilidade, uma proibição ou uma obrigação. Um material didático adequado deve apresentar o modal dentro do seu contexto real de uso para que os alunos possam compreendê-lo quando lerem os manuais. Assim, a apresentação de exemplos somente com unidades simples ou *you* como sujeitos,

[18] São considerados colocados os itens lexicais localizados em um horizonte de três casas à direita e três à esquerda, com frequência mínima de quatro e valor LL de 6,63 ou maior.

provavelmente não capacitará os alunos a entender as orações com modais nos manuais, já que não representam a maioria dos casos.

Quanto aos colocados à direita foi possível notar que alguns verbos se repetem nos dois manuais, como *be*, *cause* e *result*.[19] Como já visto, as situações que apresentam os verbos *cause* e *occur* como colocados, não expressam mera possibilidade, mas anunciam uma consequência negativa, fruto de uma ação mal realizada ou de um problema no funcionamento da aeronave. Ressalto que essa característica deveria ser explicitada em livros de inglês para aviação.

Apesar de *by* aparecer como colocado de *can* na voz passiva em ambos os subcorpora, nota-se que a grande maioria das ocorrências de *can* na voz passiva acontece sem a presença do agente que executa ou deve executar a ação expressa pelo verbo principal. Essa informação é importante porque muitos livros-texto e gramáticas da língua inglesa apresentam a voz passiva geralmente acompanhada pela preposição *by*. Entretanto, esse não é o caso no CA e tais resultados devem ser considerados nos exemplos e exercícios nesta área.

A presença de *can* + *be* deve ser destacada. Essa é a associação mais frequente entre todas as associações no CA (considerando todos os casos, voz ativa e voz passiva juntas). Apesar de *be* não ter sido considerado colocado na voz passiva, por fazer parte da estrutura, ele encontra-se também presente nesses casos. É de extrema importância que essa característica seja explicitamente apresentada nos materiais didáticos, pois a estrutura modal na voz passiva reúne os dois maiores causadores de dificuldades de compreensão de textos técnicos entre falantes não-nativos (TRIMBLE, 1985). Nesse sentido, pilotos e mecânicos devem ser ensinados a distinguir entre a estrutura modal + *be* nas vozes ativa e voz passiva para, ao lerem os manuais (de manutenção ou operação), saibam exatamente o que está sendo instruído ou indicado através do modal.

As seções *warning* e *caution* recebem um tratamento diferenciado nos dois manuais. Reproduzo um exemplo no MO para pilotos:

---

[19] Apesar de haver uma incidência muito mais alta desses verbos no MO.

“*CAUTION: An operating procedure, technique, etc., that can result in damage to equipment if not carefully followed.*”

É possível observar que a seção não é devidamente destacada, somente CAUTION está em caixa alta. Blake e Bly (1993) sugerem que destaque gráfico para as seções cruciais em um manual é fundamental para salientar a importância da seção de forma. Por outro lado, no MM, a seção de *warning* aparece da seguinte forma:

*W_A_R_N_I_N_G_: MAKE SURE EACH TUBE AND THE PORT FITTINGS HAVE TAGS TO IDENTIFY THE CORRECT INSTALLATION LOCATION. IF YOU DO NOT PUT TAGS ON THE TUBES AND PORT FITTINGS, CROSS-CONNECTION OF THE TUBES CAN OCCUR DURING INSTALLATION. IF THE TUBES ARE CROSS-CONNECTED, UNINTENDED OPERATION OR MALFUCTION OF AIRPLANE SYSTEMS CAN RESULT AND CAUSE INJURY TO PERSONS AND DAMAGE TO EQUIPMENT.*

Pode-se notar que essa seção é destacada pelo uso da caixa alta e tende a ser longa, explicando os procedimentos e as consequências mais detalhadamente.

Uma possível explicação para esse diferente tratamento pode ser atribuída às condições de trabalho e ao local de consulta dos manuais. Sem ter o propósito de avaliar o grau de profissionalização de qualquer um dos destinatários dos manuais, sejam eles mecânicos ou pilotos, é fundamental lembrar que uma ação mal realizada pelo mecânico pode colocar muitas vidas em risco, mas não necessariamente a dele, ao contrário dos pilotos, cujas ações têm reflexo na sua própria vida. Assim, as possíveis trágicas consequências de uma execução ineficaz, devem ser explicitadas de forma mais marcante no MM. Outra explicação para a diferença entre o MO e o MM, que não necessariamente anula a anterior, é a de que são necessários muitos anos de treinamento teórico e prático para que um piloto possa voar um Boeing. Dessa forma, esse profissional conhece melhor as causas e as consequências de ações e problemas com a aeronave. Por outro lado, os mecânicos fazem um curso técnico bem direcionado e mais sucinto, mas em menos de um ano já estão capacitados a trabalhar na manutenção das aeronaves.

Os resultados obtidos nesta pesquisa mostram que somente uma abordagem descritiva baseada em *corpus* pode desvendar a riqueza de

detalhes necessária para melhor compreensão de um item lexical. De acordo com Hoffmann (1998[2007]), é importante escolher e delimitar o material de ensino de forma que o aluno possa aprender, em menor tempo, a quantidade máxima possível de conhecimentos. O autor considera que esse mínimo de conteúdo necessita conter as ocorrências lexicais e gramaticais mais úteis, e que o desvendar desse mínimo constitui o cerne das pesquisas voltadas ao ensino de línguas estrangeiras com propósitos específicos.

## Referências Bibliográficas

BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. *Longman Grammar of Spoken and Written English*. London: Longman, 1999.

BYBEE, J. The Semantic Development of Past Tense Modals in English. In: Bybee, Joan; Fleischman, Suzanne. *Modality in Grammar and Discourse*. Philadelphia: John Benjamins, 503-517, 1995.

BLAKE, G. E BLY, R. W. *The Elements of Technical Writing.* The essential guide to writing clear, concise proposals, reports, manuals, letters, memos, and other documents in every technical field. New York, Macmillan General Reference, 1993.

COATES, J. *The Semantics of the Modal Auxiliaries*. London & Canberra: Croom Helm, 1983.

COATES, J. The Expression of Root and Epistemic Possibility in English. In: Bybee, Joan; Fleischman, Suzanne. *Modality in Grammar and Discourse*. Philadelphia: John Benjamins, 55-66, 1995.

DENARDIN, F. W. *Central Modals in an Operating System Corpus: Frequency, Distribution and Teaching Suggestions*. Monografia de Especialização. Porto Alegre, PUC-RS, 2008.

DENISON , D. *English Historical Syntax.* London: Longman, 1993.

FIRTH, J. R. Papers in linguistics 1934-51. London: Oxford University Press, 1957.

GABRIELATOS, C, SARMENTO, S. Central Modals in an Corpus da Aviação: Frequency and distribution. *Letras de Hoje,* Vol 41, N.2. Porto Alegre: EDIPUCRS, 2006.

HOFFMANN, L. Possibilidades de aplicação e a aplicação atual de métodos estatísticos na pesquisa de linguagens especializadas (Título Original: anwendungsmöglichkeiten und bisherige Anwengung von Statistischen Methoden in der Fachsprachenforschung, 1998). Disponível em: *Cadernos de Tradução*, Porto Alegre, nº 20, janeiro-junho, p. 61-76, 2007

HOFFMANN, L. *Conceitos Básicos da Linguística das Linguagens Especializadas* (Título Original: Grundbegriffe der Frachsprachenlinguistik, 1988). Disponível em: Cadernos de Tradução, Porto Alegre, nº 17, outubro-dezembro, p. 79-90, 2004.

HOYE, L. *Adverbs and Modality in English*. Essex: Longman,1997.

HUNSTON, S. *Corpora in Applied Linguistics*. London: Cambridge University Press, 2002.

KENNEDY, G. D. *Variation in the distribution of modal verbs in the British National Corpus*. In: Reppen, R.; Fitzmaurice, S.M., Biber, D. Using Corpora to Explore Linguistic Variation. Amsterdan: John Benjamins: 73-91, 2002.

LEECH, G. N. *O significado do verbo em inglês*. São Paulo: Ática.1989

McENERY, T; XIAO, R. e TONO, Y. *Corpus-based Language Studies: an advanced resource book*, Oxon: Routledge, 2006

NEVES, Maria Helena de Moura. *A gramática funcional*. São Paulo:

Martins Fontes, 2004.

PALMER, F.R. *Mood and Modality*. Cambridge: Cambridge University Press, 1986.

PALMER, F.R. *Modality and the English Modals*. 2nd Edition, London: Longman, 1990.

PALMER F.R. Modality in English: Theoretical, descriptive and typological issues. In; Facchinetti, Roberta; Krug, Manfred; Palmer, Frank: *A corpus based study of grammaticalization*: *Modality in Contemporary English*. Cambridge: Cambridge University Press 2003.

QUIRK, R.; GREENBAUM, S.; LEECH, G.; SVARTIK, J. *A Comprehensive Grammar of the English Language*. London: Longman, 1985.

SARMENTO, S. *Distribution of Modal Verbs in an Corpus da Aviação*. Proceedings of the Corpus Linguistics Conference, 2005a. Disponível em <http://www.corpus.bham.ac.uk/pclc> Acessado em 10 de abril de 2007.

SARMENTO, S. *A Pragmatic Account of Aviation Modals*. In: ESP-WORLD. Issue 3(11), Volume 4, 2005b. Disponível em <http://www.esp-world.info/contents.htm> Acessado em 15 de abril de 2007.

SARMENTO, S. *O uso dos verbos modais em manuais de aviação em inglês: Um estudo baseado em corpus.* Tese de doutorado. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2008.

SHAWCROSS, P. *English for Aircraft Maintenance*. Paris: Belin, 1992.

SINCLAIR, J. *Corpus, Concordance, Collocation.* Oxford: OUP, 1991.

SWALES , J.M. *Genre Analysis*. Cambridge: Cambridge University Press, 1990.

TRIMBLE, L. *English for Science and Technology. A discourse approach*. Cambridge: C.U.P., 1985.

# MACROESTRUTURA E MICROESTRUTURA DO *DICIONÁRIO DE LINGUÍSTICA DA ENUNCIAÇÃO* E AS BALIZAGENS EPISTEMOLÓGICAS DE UM CAMPO: A TEORIA DE ÉMILE BENVENISTE EM EXAME

*Valdir do Nascimento Flores*[*]

***Resumo:*** *O objetivo deste capítulo é esclarecer em que medida a apresentação das teorias feita no Dicionário de Linguística da Enunciação é responsável pela indicação de um caminho epistemológico de leitura tanto do campo enunciativo, em seu conjunto, quanto das teorias, em sua especificidade. São analisados aspectos técnicos da produção da obra, tais como o desenho da macroestrutura – principalmente, a apresentação (Palavras ao leitor) e os Dados biográficos e bibliográficos dos teóricos em foco – e a configuração de sua microestrutura –, em especial, a lista de termos e as relações entre os termos sugeridas pelo campo Termos relacionados, presente na estrutura do verbete. O caráter enciclopédico da obra é apreciado tendo em vista a sua natureza didática e a necessidade de uma adequada representação do campo de conhecimento em foco.*

## Introdução

O objetivo inicial que temos com este trabalho – implícito na primeira parte do título acima – é esclarecer ao leitor do *Dicionário de linguística da enunciação* (DLE) [1] em que medida a apresentação das

---

[*] Doutor em Linguística, professor do PPG-Letras do Instituto de Letras da UFRGS, pesquisador PQ-CNPQ. Coordenador do projeto de elaboração do *Dicionário de linguística da enunciação*.

[1] O *Dicionário de linguística da enunciação*, organizado por Leci Borges Barbisan, Maria José Bocorny Finatto, Marlene Teixeira e Valdir do Nascimento Flores e publicado pela Editora Contexto em 2009, é o resultado do trabalho de uma equipe de 40 profissionais durante mais de cinco anos. Ele reúne termos e definições de 14 teorias do campo da enunciação em mais de 400 verbetes. Usamos a sigla DLE para referir, no corpo deste texto, o *Dicionário de linguística da enunciação*. As passagens referidas são notadas com a sigla seguida da página citada.

teorias feita no dicionário é responsável pela indicação de um caminho epistemológico de leitura tanto do campo enunciativo, em seu conjunto, quanto das teorias, em sua especificidade.

Para tanto, analisamos aspectos da macroestrutura – principalmente, a apresentação (*Palavras ao leitor*) e os *Dados biográficos e bibliográficos dos teóricos em foco* – e da microestrutura do DLE –, em especial, a lista de termos e as relações entre os termos sugeridas pelo campo *Termos relacionados*, presente na estrutura do verbete.

Esperamos, com isso, poder avaliar, em linhas gerais, a necessidade e a relevância da produção de materiais de tipo enciclopédico, tal como é o DLE.

Outro objetivo que temos – explícito na segunda parte do título – é analisar como os aspectos da macro e da microestrutura se refletem na apresentação de uma teoria em particular, no caso, a teoria enunciativa de Émile Benveniste.

Para realizar os objetivos, seguimos o seguinte percurso: a) inicialmente (cf. item 1), avaliamos, ao menos em parte, a relevância do DLE no contexto da linguística brasileira. Fazemos isso a partir da análise de aspectos da macroestrutura do dicionário; b) em seguida (cf. item 2), valendo-nos da experiência de coordenação do projeto do DLE, expomos algumas das dificuldades de natureza teórica enfrentadas durante a elaboração do DLE com ênfase no tratamento da teoria de Benveniste; c) também tecemos considerações a respeito da lista de termos presente no DLE e da estrutura do verbete com destaque para o campo *Termos relacionados* (cf. item 3); d) a seguir, com nossa atenção voltada à teoria benvenistiana, fazemos pequena nota sobre a terminologia[2] do autor como forma de dar a conhecer ao leitor brasileiro o estado da arte sobre o tema (cf. item 4) e como o DLE se inclui nesse cenário; d) adiante, mostramos algumas das relações conceituais sugeridas pelo DLE, indicando como elas podem ancorar a hipótese específica de que a teoria de Benveniste sustenta-se sobre um princípio geral (cf. item 5); e) finalmente, elaboramos as conclusões gerais.

---

[2] Utilizamos duas grafias para *terminologia*: com *t* maiúsculo indicando tratar-se de uma área de estudos; com *t* minúsculo indicando o conjunto de termos ou termo de uma dada especialidade.

## 1. O *DLE* no contexto da linguística brasileira

A recente publicação do *Dicionário de linguística da enunciação* apresenta uma nova realidade, e mesmo uma nova perspectiva, à linguística brasileira: a produção de discursos de natureza enciclopédica em torno de um determinado campo epistemológico (área, disciplina, teoria etc.).

Não é praxe na linguística do Brasil a elaboração de dicionários especializados. Nos últimos anos, geralmente, temos nos limitado à tradução dos dicionários de linguística, o que não deixa de ser uma opção legítima, que não implica perda de qualidade do conhecimento da área. No entanto, a elaboração de um dicionário especializado de um determinado campo do saber impõe a quem o faz problemas diferentes, se compararmos com os problemas pertinentes ao campo da tradução e mesmo da adaptação à língua de chegada.

Produzir um dicionário exige a tomada de decisões que refletem a configuração epistemológica do campo na comunidade científica em questão. Ou seja, exige que os especialistas confirmem, ou não, que aquilo que é válido em outras comunidades também é válido, ou não, na comunidade científica na qual o dicionário está sendo produzido.

No caso do DLE, isso é perceptível tanto na macroestrutura quanto na microestrutura. Por ora, pensemos na macroestrutura.

Encontramos na macroestrutura do DLE – feito sobre um *corpus* constituído de 14 *Teorias da enunciação* –, dentre outras indicações, uma apresentação teórica bastante detalhada, denominada *Palavras ao leitor* –, com cerca de 15 páginas e incluindo uma *árvore de domínio* –, uma *Lista alfabética de termos* em que é indicado a qual autor pertence cada termo, além de *Dados biográficos e bibliográficos dos teóricos em foco* em é apresentado cada um dos teóricos cuja terminologia é contemplada no DLE.

Essas informações têm um sentido específico, considerado o contexto da linguística brasileira, e isso, ao menos, por um motivo: o campo da linguística enunciativa não tem tradição consolidada no cenário científico brasileiro ao contrário do que vemos em países como a França, por exemplo. No Brasil, as chamadas *Teorias da enunciação*, entre os anos 1970 e 1990, ou foram associadas às teorias

discursivistas, em especial a Análise de Discurso de linha francesa, ou às teorias da linguística textual, ou ainda às da pragmática. Em outras palavras: não se operou, nessa época, com uma distinção explícita entre o campo enunciativo e os outros campos de estudos do uso da linguagem. Nesse caso, um dicionário da natureza do DLE cumpre o papel de colaborar com a sistematização do conhecimento básico da área, além de compartilhá-lo de maneira didática. Com o DLE, o estudante brasileiro de linguística – usuário pretendido do DLE – passa a ter a seu alcance um conjunto de estudos especificamente voltados ao campo enunciativo.

Assim, o grande número de informações constantes na macroestrutura do DLE cumpre a função de contextualizar o conteúdo apresentado na microestrutura.

Com isso, queremos dizer que, caso o DLE fosse feito em outro país – a França, por exemplo – alguns dos itens da macroestrutura talvez fossem dispensáveis, uma vez que tal contextualização seria supérflua nesse outro cenário.

Tomemos como exemplo do que estamos dizendo o *Lexique de linguistique énonciative* produzido por M.L. Groussier e C. Rivière e publicado na França em 1996, portanto mais de uma década antes da publicação do DLE no Brasil. Após uma breve introdução de não mais que duas páginas, *Quelques mots de présentation*, onde os autores contextualizam genericamente o surgimento das *Teorias da enunciação*, dizem os autores do *Lexique*:

> Le lexique qui est présenté ici se place donc dans la mouvance d'Antoine Culioli. (...) Le but est de donner un moyen d'acèss facile à la Théorie des Opérations Énonciatives de Culioli grâce au classement alphabétique des termes, allié à de nombreux renvois (M.L. Groussier e C. Rivière, 1996, p. 6).

Ora, não deixa de causar espécie aos olhos do linguista brasileiro que, na França, se apresente, em não mais de duas páginas, um dicionário sobre a *Teoria das operações enunciativas* (TOE), sem que, em função disso, se tenha de apresentar a dita teoria e mesmo

justificar a existência do campo[3]. Isso demonstra o quão distinta é a configuração epistemológica de um campo em um dado contexto. No Brasil, por exemplo, isso seria absolutamente inviável, uma vez que são raros os trabalhos feitos com base na TOE.

Apenas a título de informação suplementar: o *Lexique de linguistique énonciative* traz em sua macroestrutura, além da já lembrada apresentação, apenas um *Index anglais* e um *Index français*, o que é muito pertinente já que, como é sabido, a teoria de Antoine Culioli – o autor da Teoria das operações enunciativas – tem larga circulação no mundo anglófono.

Isso corrobora nosso viés de leitura segundo o qual a produção de um dicionário especializado exige que se atente para o contexto no qual ele estará inserido. Seguindo essa linha de raciocínio, não seria pertinente que constasse do DLE um índice em inglês, uma vez que tal procedimento não encontraria justificativa entre nós brasileiros.

Como se pode prever, então, a tarefa de fazer um dicionário especializado não é simples e, para ser efetivada, exige articulação entre diferentes profissionais (especialistas na área a ser dicionarizada, terminólogos, tradutores, entre outros).

Muitos são os entraves para levar a cabo um empreendimento desse tipo. Falemos um pouco neles.

## 2. Das dificuldades para elaboração de um *Dicionário de Linguística da Enunciação* no Brasil

As principais dificuldades enfrentadas na elaboração do DLE foram a heterogeneidade teórica do campo em exame e a forma pouco regular de organização da terminologia de alguns dos autores integrantes do *Dicionário.*

De um lado, há no DLE autores que têm explícita preocupação de estabelecimento de uma terminologia coerente. É o caso de Oswald Ducrot, por exemplo, autor que, de maneira recorrente em seus artigos e livros, lista os termos utilizados em sua teoria, esforçando-se para defini-los de forma clara. Há autores, como A. J. Greimas, que,

---

[3] A *Teoria das operações enunciativas* é uma das 14 teorias contempladas no DLE.

inclusive, têm seu próprio dicionário[4], o que facilita sobremaneira o mapeamento de termos e definições.

De outro lado, há autores como Émile Benveniste e Mikhail Bakhtin nos quais vemos grande flutuação terminológica e de definibilidade.

Além dessas dificuldades, outras se apresentaram de maneira imperativa.

Há no campo enunciativo grande número de autores que não formularam um modelo acabado de análise da enunciação. São autores que refletiram sobre a enunciação, mas que não finalizaram uma metodologia de análise da linguagem. A teoria desses autores é, em grande medida, produto da leitura do conjunto de seus escritos. São os casos de: Michel Bréal, Charles Bally, Émile Benveniste e Mikhail Bakhtin.

Há também autores que produziram um pensamento que ultrapassa o campo da enunciação e que não podem ser integralmente reduzidos às temáticas circunscritas ao campo da enunciação. Isso exigiu dos elaboradores dos verbetes que fosse estabelecido um recorte no interior do conjunto dos trabalhos desses autores de forma a destacar do conjunto apenas o concernente ao campo enunciativo. Esse recorte se faz notar na lista de termos de cada autor. Ela reflete o entendimento dos especialistas a respeito de quais noções de cada autor são fundamentais para a reflexão sobre enunciação feita por esse autor.

Exemplos disso não faltam no DLE: o conjunto da obra de Roman Jakobson versa sobre poesia, pintura, métrica, teoria linguística, folclore, fonologia, patologia da linguagem, aquisição da linguagem, línguas no mundo, semiótica e *shifters*. A teoria enunciativa de Jakobson, no entanto, é consubstancial à noção de *shifter*[5]. Logo, os elaboradores precisaram recortar a reflexão sobre o *shifter* do conjunto dos trabalhos de Jakobson. Ou ainda: em Charles Bally, encontramos uma reflexão extensa sobre o francês falado, a didática da língua, a literatura e a estilística que é propriamente a sua teoria da enunciação, portanto, de onde emanam os termos que integram o DLE.

[4] Cf. GREIMAS, Algirdas Julien e COURTES, Joseph. *Dicionário de semiótica*. São Paulo: Contexto, 2008.

[5] Termo que Jakobson utiliza no texto *Shifters, verbal categories, the russian verb* (1957) e que é importado da obra de O. Jespersen.

Enfim, o DLE é também um espelho do entendimento que temos a respeito da episteme do campo cuja configuração, embora parcial, encontra-se representada na *Árvore de domínio* constante em *Palavras ao leitor*.

## *2.1. Um exemplo de dificuldade: a flutuação terminológica em Benveniste*

Em função dos objetivos que temos, neste item, avaliamos mais detidamente a flutuação terminológica e, como veremos, também conceitual e as dificuldades implicadas na elaboração dos verbetes referentes à teoria benvenistiana.

Em Benveniste, a flutuação se justifica, uma vez que não encontramos em seus textos um modelo acabado para a análise da linguagem. A dita *Teoria da Enunciação* de Benveniste – vale a pena observar, aqui, que tal expressão não fora efetivamente usada por Benveniste – é, na verdade, uma dedução feita pelos leitores dos artigos que estão reunidos em PLG I e PLG II[6]. Logo, podemos ver, então, que essa teoria enunciativa é mais uma construção de leitura do que propriamente uma proposição de seu autor.

Quando lemos os PLG, sem pré-conceitos sedimentados, vemos que cada texto de Benveniste contém, em si, além de teorização específica, análises que atendem a tal especificidade. É claro que percebemos um eixo comum[7] a todas as reflexões benvenistiana, que é explicitada por ele mesmo ao nomear a quinta parte dos PLG de *O homem na língua*.

---

[6] Trata-se de *Problemas de linguística geral I,* originalmente, publicado em 1966, e *Problemas de linguística geral II*, originalmente, publicado em 1974. As referências a tais livros serão feitas, aqui, conforme a edição brasileira e de acordo com o seguinte sistema: sigla, indicação do volume, seguida de indicação da página.

[7] Isso não nos impede de ver um princípio geral que norteia toda a reflexão benvenistiana no campo da enunciação. Trata-se, sempre, de estudar a *presença do homem na língua* (ele também usa *linguagem*). Sobre isso ver:
FLORES, Valdir. “Por que gosto de Benveniste? Um ensaio sobre a singularidade do homem na língua”. In: *Letras de hoje*. Porto Alegre, v.39, n.4, dez. 2004, p. 217-30.

Isso nos permite afirmar que a teoria de Benveniste é antes de tudo uma teoria das *marcas do homem na língua*, mas também na linguagem, uma vez que ele se vale da reflexão sobre as línguas para generalizar seu raciocínio.

No entanto, esse eixo comum está longe de configurar um método pronto. Eis, portanto, uma das causas de tantas diferenças conceituais entre os textos de Benveniste. Tomemos apenas um exemplo dessa flutuação conceitual: a palavra *enunciação*.

Ela aparece em diferentes textos dos *Problemas I e II* e muitas vezes em cada um desses textos e, tanto num caso quanto em outro, nem sempre com o mesmo sentido. Conclusão: há muitas definições de *enunciação* em Benveniste, embora, especialmente no Brasil, se privilegie aquela dada pelo autor em 1970 – “este colocar a língua em funcionamento por um ato individual de utilização” (PLG II, p. 82).

Aliás, a flutuação conceitual[8] é visível até mesmo dentro do mesmo texto. Somente no famoso texto de 1970 – *O aparelho formal da enunciação* – há 69 ocorrências da palavra *enunciação*, nem sempre sinônimas entre si. Observemos algumas passagens do texto: após a definição acima lembrada, encontramos: “Este grande processo pode ser estudado sob diversos aspectos. Veremos principalmente três”. (PLG II, p. 82)

A enunciação, nesse caso, é entendido como *um grande processo* (em outro momento, Benveniste dirá que a enunciação é um *ato*) e pode ser visto *sob diversos aspectos*:

a) O aspecto vocal:

“O mais imediatamente perceptível e o mais direto – embora de um modo geral não seja visto em relação ao fenômeno geral da

---

[8] Além da flutuação conceitual, há também uma flutuação terminológica. Por exemplo, Benveniste parece, em alguns textos, dar nomes diferentes para o mesmo conceito. Observemos as passagens seguintes dentro do mesmo texto de 1956, *A natureza dos pronomes*: “o emprego tem como condição a **situação de discurso** e nenhuma outra” (PLG I, p. 280) e “*eu* só pode ser identificado pela **instância de discurso** que o contém e somente por aí.” (PLG I, p. 278). Não seria absurdo pensar que aqui **instância de discurso** e **situação de discurso** são designações diferentes para um mesmo conceito.

enunciação – **é a realização vocal da língua**.” (PLG II, p. 82) [grifos nossos];

“Na prática científica procura-se eliminar ou atenuar os traços individuais da **enunciação fônica** recorrendo a sujeitos diferentes e multiplicando os registros, de modo a obter uma imagem média de sons, distintos ou ligados. Mas cada um sabe que, para o mesmo sujeito, os mesmos sons não são jamais reproduzidos exatamente, e que a noção de identidade não é senão aproximativa mesmo quando a experiência é repetida em detalhe. Estas diferenças dizem respeito à diversidade das situações nas quais a **enunciação é produzida**.” (PLG II, p. 82-83) [grifos nossos];

b) O aspecto da semantização:

“O mecanismo desta produção é um outro aspecto maior do mesmo problema. **A enunciação supõe a conversão individual da língua em discurso**.” (PLG II, p. 83) [grifos nossos];

“É a **semantização** da língua que está no centro **deste aspecto da enunciação**, e ela conduz à teoria do signo e à análise da significância.” (PLG II, p. 83) [grifos nossos];

c) O quadro formal de realização:

“Pode-se, enfim, considerar uma outra abordagem, que consistiria em definir a **enunciação no quadro formal de sua realização**. É o objeto próprio destas páginas.” (PLG II, p. 83) [grifos nossos].

Considerando-se o que foi dito até aqui, podemos ver que, do ponto de vista da elaboração do DLE, algumas questões se impõem:

1) Como operacionalizar em um verbete casos em que se tem a mesma designação – *enunciação* e nota 8 (supra), por exemplo – e definições nem sempre coincidentes no conjunto da obra que serve como *corpus* de referência?
2) Também: como não dar a falsa impressão de que a teoria é uma mera justaposição de termos? Ou seja: como deixar claro ao leitor que se trata de um sistema conceitual?

Para responder a tais questões vale deter-se um pouco mais sobre a microestrutura do DLE.

## 3. A microestrutura do DLE: um princípio epistemológico de leitura

O DLE parte de um princípio epistemológico geral de leitura da Linguística da Enunciação: as teorias podem ser apresentadas como um conjunto de conceitos sistematicamente relacionados entre si. Tais conceitos são identificados pelos termos que os nomeiam, o que permite responder a segunda questão posta no item anterior. Não se trata, então, de mera justaposição de termos mas da explicitação de um sistema de conceitos.

Assim, podemos dizer que a teoria de Benveniste aceita ser lida como uma complexa rede conceitual cujos termos e noções estão interligados a partir de diferentes relações.

No DLE, essas relações são contempladas no verbete no campo *Termos relacionados*, que é, por sua vez, limitado a um grupo de no máximo três termos, em função dos fins didáticos do DLE.

Observemos a estrutura do verbete do DLE, aqui sintetizada a partir do que é apresentado no "Guia do Usuário" (cf. DLE: p. 31-32):

1- Termo: expressão relevante coletada na(s) obra(s) do autor. O número ao lado do termo indica que a mesma forma, com sentido diferente, ocorre em outros autores. Classificação gramatical do termo ou expressão (s.f. = substantivo feminino).
2- Classificação gramatical do termo
3- Nome: autor/teórico em foco.
4- Outras denominações: variante do termo verificada na obra do mesmo autor.
5- Definição: coletada na(s) obra(s) do autor e/ou elaborada pela equipe responsável.
6- Fonte da definição: indica a fonte consultada conforme codificação
7- Nota explicativa: informações que complementam a definição.
8- Fonte da nota: indica a fonte consultada conforme codificação.
9- Leitura recomendada: fontes que podem complementar as informações oferecidas sobre o termo.
10- Termos relacionados: outros termos presentes no dicionário (no máximo três) que, em conjunto com o termo em foco, integram uma rede de noções afins de um mesmo autor. Sua consulta amplia a compreensão da noção em foco no verbete. Todos os itens indicados como "termos relacionados" correspondem a verbetes com informações completas.

Quadro 1 – síntese da estrutura do verbete do DLE

A estrutura do verbete do DLE, acima sintetizada, permite ver que há uma espécie de roteiro de leitura da teoria em foco. Isso pode ser visualizado no DLE na medida em que percebemos que os termos e as noções integrantes de um dado conceito contêm outros termos e noções e estes, por sua vez, estão contidos em muitos outros.

Vejamos, no quadro abaixo, a título de exemplo, como o que dissemos se aplica a um verbete da teoria de Benveniste no DLE:

**"**Correferência *s. f.* Benveniste.
**Outras denominações: co-referir.**
**Definição: possibilidade linguística própria do co-locutor de partilhar da referência do discurso do locutor.**
**Fonte da definição:** BEN89**: 84.**
**Nota explicativa: A co-referência é fundante do diálogo e está estreitamente ligada à ideia da intersubjetividade. Por exemplo, o locutor, ao enunciar "ontem...", instaura a possibilidade do co-locutor de atribuir a "ontem..." o mesmo sentido dado pelo locutor.**
**Fonte da nota:** EBE06.
**Leitura recomendada:** BEN89C.
**Termos relacionados:** intersubjetividade**,** pessoa**,** referência**." (DLE: p. 70)**

Quadro 2 – exemplo de um verbete da teoria de Benveniste no DLE

Como é possível notar, o campo *Termos relacionados* indica ao leitor que a definição do termo "Correferência" pode ser expandida se forem consultados os termos *intersubjetividade, pessoa* e *referência*.

Evidentemente, essa indicação não explicita a natureza das relações (hierárquicas, paralelas, transversais etc.) que são supostas entre os termos, informação esta desnecessária para um dicionário do padrão do DLE. No entanto, é afirmado: há relações.

Admitido este raciocínio, pode-se ver que há em Benveniste conceitos, termos e noções cuja compreensão decorre das relações que estes mantêm com outros conceitos, termos e noções. Isso significa dizer que é difícil, nessa teoria, estudar-se um elemento isoladamente.

Certamente, as relações variam na medida em que variam os termos que servem como ponto de partida. Poderíamos inclusive questionar a pertinência do que é apresentado como termo relacionado. Isso é de menor importância para o que estamos

explicando. Na verdade, defendemos a existência das relações sem, no entanto, defendermos que elas se configurem dessa ou daquela forma. É secundário, ao menos neste momento, o fato de se aceitar, ou não, qual termo integra cada conjunto de relações. O essencial é perceber que a reflexão benvenistiana não pode ser lida de maneira linear.

Ainda na mesma direção, cabe prestar atenção nos campo *Definição* e *Nota explicativa*. É a partir deles que vislumbramos respostas à primeira questão antes apresentada (cf. supra).

O campo *Definição* contém a acepção do termo devido à indicação de suas características genéricas e específicas. Prioriza-se a acepção mais representativa do termo. Em *Nota explicativa*, complementa-se a definição dada através de explicações, exemplos, explicitação de nuances conceituais, finalidade e forma de inclusão do termo na teoria em foco.

A partir disso, a seguir fazemos ainda algum comentário sobre a escolha dos termos. Para tanto, como forma de exemplificação, estaremos circunscrito, no item seguinte, à terminologia da teoria enunciativa de Émile Benveniste.

## 4. Nota sobre a terminologia de Benveniste

No Brasil, não temos tradição de estudos terminológicos de autores do campo da linguística. No entanto, no cenário da linguística geral, isso é mais comum. O estudo da terminologia de Benveniste, por exemplo, não é recente. Muito já se produziu sobre isso.

Em primeiro lugar, há uma espécie de léxico técnico presente em forma de *Índice remissivo* em apêndice aos PLG I e II. Como se sabe, a publicação do primeiro volume é datada de 1966 e a do segundo volume é de 1974. A organização do PLG I é assinada por Benveniste; a do PLG II por M. Dj. Moinfar[9]. Como bem é explicado no *Prefácio* dos livros, tanto em um caso quanto em outro, tem-se a supervisão direta de Benveniste.

---

[9] Benveniste morreu em 1976.

A partir disso, pode-se concluir que muitas das palavras que estão nos *Índices remissivos* dos dois volumes se não têm estatuto de termo da teoria benvenistiana têm, ao menos, lugar de destaque na reflexão do autor. De qualquer maneira, há nos *Índices* alguma indicação de um caminho a seguir quando o que está em questão é a terminologia do autor.

Mas o primeiro grande trabalho que realmente aborda a questão terminológica em Benveniste é um *Lexique* escrito por Jean Claude Coquet e Marc Derycke intitulado *Le lexique d'E. Benveniste*. Escrito em dois volumes e publicado em 1971 e 1972[10], respectivamente, é formado por mais de 180 termos que dizem respeito a toda a reflexão de Benveniste, não apenas a circunscrita ao campo da enunciação. O trabalho de J-C Coquet e M. Dericke é consulta obrigatória a todos os que, de um modo geral, se interessam pelo pensamento de Benveniste.

Em 1975 vem a público um exaustivo trabalho de levantamento bibliográfico escrito por Mohammad Djafar Moïnfar[11]. Trata-se de uma "bibliografia que tem por objetivo estabelecer uma lista tão completa quanto possível" (Moïnfar, 1975, p. IX) segundo o plano seguinte[12]: obras (classificadas em ordem cronológica), artigos (classificados cronologicamente e subclassificados segundo o assunto), índice de assuntos, resenhas e comunicações feitas na *Société de Linguistique*. De certa forma, nesse trabalho, temos uma indicação terminológica, em especial, no *Índice de assuntos*.

Além desses trabalhos, a partir dos anos oitenta, Claudine Normand também publica textos voltados à questão terminológica de

---

[10] Cf. COQUET, J-C; DERYCKE, M. *Le lexique d'E. Benveniste* (vol. I). Documents de travail et pré-publications, nº 8, serie A, Centro Internazionale di Semiótica e di Linguistica, Università di Urbino, 1971, p.1-40.
COQUET, J-C; DERYCKE, M. *Le lexique d'E. Benveniste* (vol. II). Documents de travail et pré-publications, nº 8, serie A, Centro Internazionale di Semiótica e di Linguistica, Università di Urbino, 1972, p.41-78.

[11] DJAFAR, Moïnfar Mohammad. *Bibliographie des travaux d'Emile Benveniste*. In : *Mélanges linguistiques offerts à Émile Benveniste*, Louvain, Peeters. (Collection Linguistique publiée par la Société de Linguistique de Paris LXX).

[12] Ver também: DJAFAR, Moïnfar Mohammad. « Sur la terminologie de Benveniste ». In : *LINX*, numero special, 1997, p. 365-373.

Benveniste. Tais textos são condizentes com uma abordagem investigativa que busca vislumbrar oscilações de sentido no tratamento conceitual e terminológico dado por Benveniste. Os artigos *Les termes de l'énonciation de Benveniste* (1986), *Constitution de la sémiologie chez Benveniste* (1989), *Émile Benveniste: quelle sémantique?* (1996) e *Sémiologie, sémiotique, sémantique: remarques sur l'emploi de ces termes par Émile Benveniste*, entre outros, fazem explicitamente referência ao tratamento terminológico da obra benvenistiana.

Recentemente, encontramos um livro de Aya Ono – cujo título inspirador, *La notion d'énonciation chez Émile Benveniste*, já indica a problemática terminológica, "la notion", – que expõe com rigor e complexidade uma infinidade de relações conceituais e noções associadas à noção de enunciação na obra de Benveniste. Desse mesmo tipo é o trabalho de Gérard Dessons, *Émile Benveniste, l'invention du discours*, onde encontramos um detalhado mapeamento de noções como *comunicação, discurso, língua, linguagem, significação, subjetividade, tempo* e *sujeito*.

Também em *Sujet de l'énonciation et ébauche d'une réflexion sur la singularité enonciative* (cf. Flores, no prelo) fazemos uma investigação terminológica na obra destacando, principalmente, as relações conceituais entre *língua, linguagem, línguas, homem, locutor, sujeito, intersubjetividade, subjetividade* e *pessoa.*

Finalmente, há a terminologia presente no DLE. Sobre ela, iremos nos deter no item seguinte.

## 5. As relações conceituais: o exemplo da teoria benvenistiana

O DLE propõe uma lista de termos e uma organização conceitual para a teoria enunciativa de Émile Benveniste. A seguir, apresentamos a lista de termos, na primeira coluna, acompanhada dos termos relacionados, na segunda coluna.

| COLUNA A | COLUNA B |
|---|---|
| **1. Agenciamento -** | apropriação, referência, sintagmatização |
| **2. Aparelho formal da enunciação-** | enunciação, língua, discurso. |
| **3. Apropriação -** | atualização, língua, subjetividade. |
| **4. Atualização –** | apropriação, enunciação, língua-discurso |
| **5. Compreensão -** | frase, palavra, reconhecimento. |
| **6. co-referência-** | intersubjetividade, pessoa, referência. |
| 7. **Correlação de pessoalidade-** | correlação de subjetividade, não-pessoa, **pessoa** |
| 8. **Correlação de personalidade** | correlação de subjetividade, não-pessoa, **pessoa** |
| 9. **Correlação de subjetividade -** | correlação de pessoalidade, não-pessoa, **pessoa** |
| 10. **Dêixis -** | indicadores de subjetividade, referência |
| 11. **Diálogo -** | aparelho formal da enunciação, intersubjetividade, pessoa. |
| 12. **Discurso -** | enunciação, língua-discurso, semântica |
| **13. Domínio do semiótico -** | forma, língua, semântico |
| 14. **Ele -** | correlação de pessoalidade, correlação de subjetividade, pessoa |
| 15. **Enunciação -** | aparelho formal da enunciação, língua-discurso, semântico. |
| **16. Enunciado -** | atualização, enunciação, palavra. |

| | |
|---|---|
| **17. Eu -** | intersubjetividade, locutor, semântico |
| **18. Forma -** | semântico, semiótico |
| **19. Forma vazia -** | discurso, signo, subjetividade |
| **20. Frase -** | enunciado, ideia, semântico |
| **21. Ideia -** | frase, referência, sentido |
| **22. Indicadores auto-referenciais -** | **indicadores de dêixis** |
| **23. Indicadores de dêixis -** | **indicadores auto-referenciais** |
| **24. Indicadores de subjetividade-** | **Instância do discurso** |
| 25. **Instância de discurso-** | enunciação, indicadores de subjetividade, semântico |
| **26. Intersubjetividade -** | enunciação, pessoa, subjetividade |
| **27. Língua –** | **Intersubjetividade, língua discurso** |
| **28. Língua-discurso -** | semântico, semiótico, sentido |
| **29. Linguagem -** | **língua**, língua-discurso |
| **30. Locutor -** | **subjetividade, sujeito** |
| **31. Metassemântica-** | enunciação, semântico, semiótico**.** |
| **32. Modo semântico -** | língua-discurso, semiótico, sentido |
| **33. Modo semiótico -** | forma, língua, semântico |
| 34. **Monólogo -** | aparelho formal da enunciação, diálogo,**pessoa** |
| 35. **Não-pessoa -** | correlação de pessoalidade, correlação de subjetividade, pessoa |
| **36. Ordem semiótica -** | forma, língua, semântico |
| **37. Palavra-** | atualização, sentido |
| **38. Pessoa -** | **eu,** sujeito, tu |
| 39. **Presente-** | intersubjetividade, tempo linguístico**.** |

| | |
|---|---|
| **40. Reconhecimento -** | atualização, compreensão, palavra**.** |
| 41. **Referência -** | língua, pessoa, situação de discurso**.** |
| **42. Semântico -** | língua-discurso, semiótico, sentido |
| **43. Semantização -** | ideia, língua-discurso, **]** semântico |
| **44. Semiótico -** | forma, língua, semântico |
| 45. **Sentido -** | discurso, forma, língua-discurso |
| **46. Signo -** | dêixis, língua, signo vazio |
| **47. Signo vazio -** | discurso, signo, subjetividade |
| **48. Sintagmatização -** | discurso, frase, palavra |
| **49. Situação de discurso -** | discurso, sujeito |
| **50. Subjetividade-** | intersubjetividade, sujeito, pessoa |
| 51. **Sui-referencial -** | indicadores de subjetividade, signo vazio, situação de discurso |
| **52. Sujeito -** | intersubjetividade, locutor, semântico**.** |
| **53. Tempo crônico -** | tempo linguístico |
| **54. Tempo da língua -** | enunciação, presente, tempo crônico |
| **55. Tempo linguístico -** | enunciação, presente, tempo crônico**.** |
| 56. **Terceira pessoa -** | correlação de pessoalidade, correlação de subjetividade, pessoa |
| **57. Unicidade -** | enunciação, pessoa |
| 58. **Verbos delocutivos -** | aparelho formal de enunciação, enunciação, subjetividade |

Quadro 3 – lista de termos acompanhada dos termos relacionados da teoria de Benveniste no DLE

Essa lista de termos tem peculiaridades que são de extremo interesse a todos os que se interessam pela teoria enunciativa de Benveniste e pela elaboração de dicionários. Vejamos algumas.

O primeiro ponto que chama a atenção é que não é fornecida a natureza das relações postas em *Termos relacionados.* Por exemplo: embora o DLE nos diga que *Não-pessoa* relaciona-se com *correlação de pessoalidade*, *correlação de subjetividade* e *pessoa*, ele não informa se tais termos têm alguma relação de hierarquia, de paralelismo, de transversalidade entre si, se são hiperônimos ou se são antônimos. O campo *Termos relacionados* apenas indica que o consulente complementará sua consulta se também contatar com os outros verbetes.

Evidentemente, essa constatação não encerra uma crítica ao DLE. A explicitação dessas informações transcende os objetivos a partir dos quais o DLE fora formulado e está mais próxima de um trabalho que visasse à apresentação de um mapa conceitual da teoria.

Assim, o campo *Termos relacionados* cumpre um papel no DLE: o de alertar para o fato de as teorias da enunciação, em geral, e a teoria de Benveniste, no caso aqui destacado, aceitar serem lidas como uma complexa rede cujos termos e noções estão interligados a partir de diferentes relações – hierárquicas, paralelas, transversais etc. – entre si.

A consequência disso é que muitos dos conceitos propostos por Benveniste, por exemplo, têm valor primitivo, na medida em que integram outros conceitos. Ou seja, os termos e as noções que fazem parte de um dado conceito contêm outros termos e noções e estes, por sua vez, estão contidos em muitos outros.

Em suma: o DLE imprime um viés de leitura da teoria enunciativa de Benveniste (e das outras também): considerá-la como uma rede de primitivos teóricos.

Certamente, as relações variam na medida em que variam os termos que servem como ponto de partida. Podemos, sem dúvida, questionar a pertinência do que é apresentado como *Termos relacionados*. Não é disso que estamos tratando. O DLE apenas explicita a existência das relações sem, no entanto, defender que elas se esgotem na maneira como estão colocadas em cada verbete. Com certeza, o mapeamento conceitual de cada termo – bem como a posição que ocupa em relação aos demais termos – é limitado uma vez

que, como dissemos, o DLE apenas admite três *Termos relacionados* para cada termo.

Disso decorre um segundo ponto de fundamental importância: o DLE, ao contrário do que poderia se esperar de um dicionário especializado, não fecha as relações de sentido numa imanência que ignoraria a epistemologia do campo. Vejamos como isso é operador da leitura que podemos fazer da teoria de Benveniste.

O termo “enunciação” é *Termo relacionado* de 11 outros termos. Considerando-se um universo de 58 termos, conclui-se que, segundo o DLE, a compreensão de cerca de 20% dos termos da teoria enunciativa de Émile Benveniste está na dependência direta do termo “enunciação”. Disso, conclui-se também que esse termo ocupa lugar de destaque na teoria.

Observemos mais de perto as implicações que uma informação dessa natureza tem relativamente à leitura de uma teoria como a de Benveniste. Considere-se apenas uma ocorrência de “enunciação” como *Termo relacionado*: no verbete “Intersubjetividade”.

Em “intersubjetividade”, encontramos remissão, além de a “enunciação”, também a “pessoa” e a “subjetividade”. Como o DLE não explicita a natureza das relações entre os termos, é aos campos *Definição* e *Nota explicativa* que devemos recorrer para melhor entendimento das relações.

Na *Definição*, encontramos sobre “intersubjetividade”: “inter-relação constitutiva da **enunciação** que pressupõe o eu e o outro mutuamente implicados” (DLE, p. 146) [grifo nosso]. Logo a seguir, em *Nota explicativa*, há: “O tema da intersubjetividade é recorrente em Benveniste, porém, o uso da palavra intersubjetividade é menos comum se comparado à **subjetividade** e à **pessoa**.” (Idem) [grifo nosso].

De imediato encontramos distribuídos entre a *Definição* e a *Nota explicativa* os três termos que constam de *Termos relacionados* e a natureza das relações entre eles começa a se explicitar.

Diz, ainda, o DLE: “intersubjetividade é apresentada por Benveniste como uma ‘condição’ da experiência humana inerente à linguagem. Essa experiência se reflete na língua” (Idem). E conclui:

> Em suma, **a Teoria da Enunciação de Benveniste tem como fundamento a noção de intersubjetividade**, já que é na e pela linguagem que o homem se constitui como sujeito, e essa condição está na dependência da existência do outro. Como exemplo, pode-se considerar que a intersubjetividade se marca na língua pela relação de oposição pessoa eu/ pessoa não-eu, distinção necessária à noção de unicidade de cada uma das pessoas. Essa unicidade se apresenta em uma relação complementar, que é intrínseca à relação de subjetividade - há pessoa subjetiva porque há pessoa não subjetiva, relação que pode inverter-se. **Intersubjetividade é, então, noção primeira, pressuposta, epistemologicamente, pela noção de subjetividade**. (Idem) [grifos nossos].

Ora, a partir disso podemos compreender que “intersubjetividade” é termo de maior importância uma vez que é “noção primeira, pressuposta epistemologicamente”. Numa tentativa de explicitar os tipos de relações possíveis entre os termos e as noções que eles designam, teríamos que reservar lugar de maior destaque a “intersubjetividade” já que tudo indica que dele decorrem muitas outras noções devido o caráter “fundante” que tem.

Assim uma continuidade possível para o DLE seria a de ver, para cada uma das teorias ali contempladas, como se organizam os termos e as definições entre si. Que relações elas estabelecem umas em relação às outras? Que primitivos teóricos cada autor elege para o seu modo de pensar? Enfim, que conhecimento sobre a linguagem cada uma supõe (ou não)? Isso ainda está por ser feito e, conforme pensamos, pode dar origem a um tipo de dicionário que objetive explicitar o sistema conceitual de um pensamento.

## Conclusão

Em termos de conclusão, cabe finalizar indicando as perspectivas que o DLE traz para a consolidação de um discurso epistemológico para a linguística brasileira.

Um dos principais indicadores que o DLE produz, na atualidade, no contexto da linguística brasileira, é o da possibilidade de construir um saber que, além de determinar criticamente os princípios, as hipóteses e os objetos construídos, também ajuda a compreender como esse campo se configurou dentro de uma organização sócio-política. Muito há ainda para ser feito em outras tantas áreas.

## Referências bilbiográficas


BENVENISTE, E. *Problèmes de linguistique générale, 1*. Paris: Gallimard, 1966.

BENVENISTE, E. *Problèmes de linguistique générale, 2*. Paris: Gallimard, 1974.

BENVENISTE, E. *Problemas de linguística geral, 1*. São Paulo: Campinas, 1988.

BENVENISTE, E. *Problemas de linguística geral, 2*. São Paulo: Campinas, 1989.

COQUET, J-C; DERYCKE, M. *Le lexique d'E. Benveniste* (vol. I). Documents de travail et pré-publications, nº 8, serie A, Centro Internazionale di Semiótica e di Linguistica, Università di Urbino, 1971, p.1-40.

COQUET, J-C; DERYCKE, M. *Le lexique d'E. Benveniste* (vol. II). Documents de travail et pré-publications, nº 8, serie A, Centro Internazionale di Semiótica e di Linguistica, Università di Urbino, 1972, p.41-78.

DESSON, Gérard. *Émile Benveniste, l'invention du discours*. Paris : Éditions IN PRESS, 2006.

DJAFAR, Moïnfar Mohammad. *Bibliographie des travaux d'Emile Benveniste*. In : *Mélanges linguistiques offerts à Émile Benveniste*, Louvain, Peeters. (Collection Linguistique publiée par la Société de Linguistique de Paris LXX).

DJAFAR, Moïnfar Mohammad. « Sur la terminologie de Benveniste ». In: *LINX*, numero special, 1997, p. 365-373.

FLORES Valdir do Nascimento; BARBISAN Leci Borges; FINATTO Maria José Bocorny; TEIXEIRA, Marlene. *Dicionário de linguística da enunciação*. São Paulo: Contexto, 2009.

FLORES, Valdir do Nascimento. “Por que gosto de Benveniste? Um ensaio sobre a singularidade do homem na língua”. In: *Letras de hoje*. Porto Alegre, v.39, n.4, dez. 2004, p. 217-30.

______. *Sujet de l’énonciation et ébauche d’une réflexion sur la singularité enonciative*. In : NORMAND, Claudine (org.) .*Paralleles floues: vers une théorie du langage.* (no prelo).

GREIMAS, Algirdas Julien e COURTES, Joseph. *Dicionário de semiótica*. São Paulo: Contexto, 2008.

GROUSSIER, M.L. e RIVIÈRE, C. *Les mots de la linguistique : lexique de linguistique énonciative*. Paris: OPHRYS, 1996.

NORMAND, Claudine. *Les termes de l’énonciation de Benveniste*. In: *Histoire Épistémologie Langage*, VIII-2, 1986.

______. *Constitution de la sémiologie chez Benveniste*. In: *Histoire Épistémologie Langage*, XI-2, 1989.

______. *Émile Benveniste : quelle sémantique?*. In: *LINX*, número especial, 1996.

______. *Sémiologie, sémiotique, sémantique : remarques sur l’emploi de ces termes par Émile Benveniste*. In: *LINX*, nº44, 2001.

ONO, Aya. *La notion dénonciation chez Émile Benveniste*. Limoges : Lambert-Lucas, 2007.